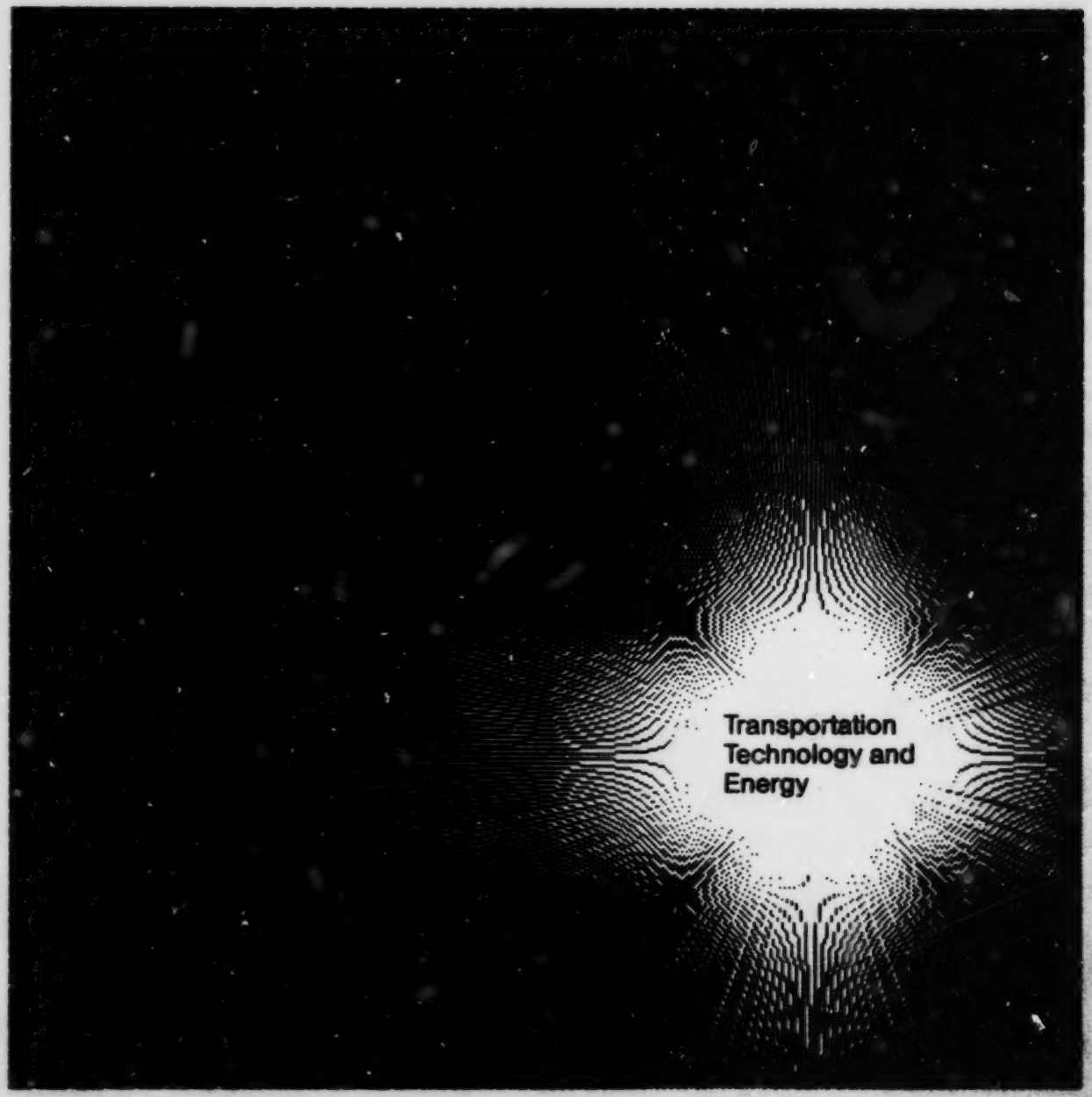
HI-84-01

Ontario

Ministry of
Transportation and
Communications

Transportation
Technology and
Energy Branch

Driver Response To Visual Highway Information Systems – Phase 1



84- 4973

Technical Report Documentation Page

MTC Report Number HI-84-01	ISBN Number:	Other Document Number:	Date: Aug 1984

Title: Driver Response to Visual Highway Information Systems - Phase I

Authors: A. Smiley, Principal Consultant
R. Dewar, Associate Consultant

MMS Man-Machine Systems Consultants Inc.

Published by:

The Transportation Technology and Energy Branch, MTC

Client Groups:

Human and Social Factors Office
and
Traffic Management and Engineering Office, MTC

Participating Agencies:

Abstract:

This report is a literature review which examines various testing methods used to evaluate visual highway information systems. These systems include static signs, both symbolic and word signs, dynamic signs, lane markings, and delineators such as pavement edge lines, reflective posts, and construction panels. The aspects of driver response examined by these testing methods were legibility distance, comprehension, processing time, conspicuity and driver-car response. Recommendations are made on general considerations in evaluating highway information systems, on specific issues in each aspect of driver response, and on the development of testing facilities.

Key Words: traffic information systems, signs, legibility, conspicuity, pavement markings

Comments:

Distribution: On request

Copyright Status: Crown copyright

Hi-84-01

Driver Response To Visual Highway Information Systems – Phase 1

A. Smiley
Principal Consultant

R. Dewar
Associate Consultant

MMS Man-Machine Systems
Consultants Inc.

Project Manager:
D. Duncan
Human and Social Factors Office, MTC

Prepared on behalf of:
Traffic Management and Engineering Office, MTC

Published by:
The Transportation Technology and Energy Branch
Ontario Ministry of Transportation and Communications
Hon. James W. Snow, Minister
H.F. Gilbert, Deputy Minister

Published without prejudice
as to the application of the findings.
Crown copyright reserved; however, this
document may be reproduced for non-commercial
purposes with attribution to the Ministry.

For additional copies write:
The Editor, Technical Publications
Ministry of Transportation and Communications
1201 Wilson Avenue
Downsview, Ontario
M3M 1J8

October 1984

Table of Contents

ACKNOWLEDGEMENTS iv

EXECUTIVE SUMMARY v

Chapter 1 : INTRODUCTION 2

Chapter 2 : LITERATURE REVIEW OF EVALUATION TECHNIQUES 5

Sec. I : EVALUATION OF STATIC SIGNS 5

LEGIBILITY DISTANCE 5
Changes in Sign Size 5
Moving Subjects Towards Signs 6
Dynamic Presentation of Signs 6
Legibility and Driver Processing Rate 6
Legibility Distance Under Degraded Visual Conditions 7

COMPREHENSION 7
Recognition Tests 7
Rating Measures 9
Semantic Differential 9
Use of Multiple Measures for Testing Comprehension 11

PROCESSING TIME 12
Processing Time Under Degraded Visual Conditions 14

CONSPICUITY 15

DRIVER-CAR RESPONSE 17

Sec. II : EVALUATION OF DYNAMIC SIGNS 18

INTRODUCTION 18

DRIVER ATTITUDE 18
Driver Attitude about Potential Use of Changeable Message Signs 18
Driver Attitude About Signs in Use 19

COMPREHENSION 21

LEGIBILITY DISTANCE 22

CONSPICUITY 24

DRIVER-CAR RESPONSE 24

Sec. III: EVALUATION OF LANE MARKINGS 28

INTRODUCTION 28

LEGIBILITY DISTANCE 28

COMPREHENSION AND PROCESSING RATE 28

DRIVER-CAR RESPONSE 30

Sec. IV: EVALUATION OF DELINEATORS 32

INTRODUCTION 32

LEGIBILITY DISTANCE 32

COMPREHENSION 32

PROCESSING TIME AND CONSPICUITY 33

DRIVER-CAR RESPONSE 34

CHAPTER 3: RECOMMENDATIONS 36

GENERAL CONSIDERATIONS IN EVALUATING HIGHWAY INFORMATION SYSTEMS 36
- Sampling of the Driver Population 36
- Assessment Under Degraded Conditions 37
- Use of Loading Tasks 37
- Dynamic vs. Static Assessment 37
- Setting of Minimum Standards 38

EVALUATION 38

LEGIBILITY DISTANCE TESTING 40
- Static and Dynamic Signs 40
- Lane Markings and Delineators 41
- Special Considerations 41

COMPREHENSION TESTING 41
- Static and Dynamic Signs 41
- Lane Markings and Delineators 42
- Special Considerations 42

PROCESSING TIME TESTING 43

CONSPICUITY TESTING 44
Static Signs 44
Dynamic Signs 44
Lane Markings and Delineators 44

DRIVER-CAR RESPONSE TESTING 45
Static Signs 45
Dynamic Signs 45
Lane Markings and Delineators 47

TEST FACILITY 47

REFERENCES 49

List of Tables and Appendices

Table I : Number of References 4
Table ii : Recommended Evaluation Techniques 39

Appendix A: Annotated Bibliography
Appendix B: Evaluation Technique Definitions
Appendix C: Statistical Considerations

ACKNOWLEDGEMENTS

The authors thank Ms. Carolyn MacGregor for writing the annotated bibliography and Dr. John Senders for valuable suggestions particularly in regard to statistical considerations.

EXECUTIVE SUMMARY

Introduction

This report discusses the evaluation of driver response to highway visual information systems. These systems were divided into four categories : (1) static signs, including both symbolic and word signs; (2) dynamic signs, including changeable message and pictogram signs as well as flashing arrow signs; (3) lane markings including the solid and dashed lines used to mark lanes and (4) delineators, which included pavement markings, roadside reflective posts and panels and barricades used around construction sites. Five aspects of driver response were identified in the process of searching the literature. These are legibility distance, comprehension, processing time, conspicuity and driver-car response. Legibility distance is the distance at which a driver can read a sign. Comprehension refers to the driver's understanding of the sign. Processing time is the time taken by the driver to understand a sign. Conspicuity is a measure of the attention-getting nature of the sign. Driver-car response refers to movements of the car initiated by a driver in response to a sign or lane marking e.g. changing lanes, braking, exiting from a freeway.

Legibility Distance Testing

Legibility distance of signs has been evaluated mainly by using slides in a laboratory or by using reduced or full size signs in the field. Techniques for measuring legibility distance involve changes in the size of the sign stimulus in one of the following ways: changing the size of the sign in order to simulate different distances, moving the subject towards the sign, or having the subject view the sign dynamically either from a car or using a film taken from a car. The first and third techniques have been used in combination: a reduced size sign is used in a field setting so that car speeds and travel distances can be reduced for reasons of economy of time as well as space.

For dynamic signs, legibility distance has been evaluated in the laboratory by using a CRT presentation of letters on a dot matrix display with the subject at a number of viewing distances, and in the field by using prototype pictogram signs and driving the subjects towards them.

Because of the dynamic nature of the driving task a dynamic test of legibility distance was advised, which would eliminate the need for additional testing of driver processing rate. It was recommended that legibility distance be tested either in the laboratory, using graphics software

and a CRT to zoom the sign towards the subject at the rate he or she would normally approach the sign, or in the field, using reduced size signs which subjects drive towards.

No recommendations could be made on legibility distance testing for lane markers or delineators because essentially no literature was found in this area.

Comprehension Testing

Techniques used to evaluate comprehension of static signs included recognition, recall, multiple choice, matching, semantic differentials, and ratings of associative strength, preference and clarity of meaning. It was generally agreed that more than one technique was necessary to fully assess comprehension of a sign.

Little literature was found on comprehension testing of changeable messages. Most studies dealt with drivers' attitudes towards or preferences for various sign messages. Understanding per se was not measured. Comprehension of symbolic dynamic signs was tested using film clips of flashing arrow signs with a multiple choice test of recognition, as well as by having subjects view prototype signs in the field and then write out the meaning of the sign.

Comprehension of lane markings was evaluated by showing subjects film clips and asking them to explain the meaning of the markings (recognition method), and view slides taken of curves at night with various lane markings and delineators and asking them to make judgements of road geometry.

It was recommended that comprehension of static and dynamic signs be assessed by first preference ranking according to which symbols or messages communicated the intended meaning best. From the preference ranking three symbols or messages should then be selected for further evaluation using the recognition test for verbal messages.

The recommendation for assessing comprehension of lane markings and delineators was to use short filmed sequences interspersed with multiple choice questionnaires asking the subject which lanes he or she was permitted to travel in.

Processing Time Testing

Processing time for static signs was tested by presenting subjects with slides of the signs and measuring time for the subjects to indicate they understood the sign (reaction time method), or by presenting slides for a

fraction of a second and measuring the percentage of correct responses (glance legibility method). No studies were found which evaluated driver processing rate for dynamic signs. Processing time for lane markings and delineators was measured using the reaction time method.

By evaluating static signs dynamically the need for processing rate testing is eliminated. It was recommended that processing rate be evaluated using reaction time measures in cases where changeable message signs were very long and for lane markings and delineators which were discontinuous, e.g. at intersections, lane drop-offs etc.

Conspicuity Testing

Conspicuity of static signs has been tested by presenting subjects with slides of signs, embedded in different backgrounds, for a fraction of a second, and asking them to report all features seen in the slides, as well as by asking subjects to judge the primary target feature and its background with respect to various attributes such as target edge contrast and sharpness etc. Conspicuity of dynamic signs has only been assessed in the simplest way possible, by simply sending out a questionnaire to drivers known to have passed a particular sign, asking them if they noticed the sign. No studies were found which assessed conspicuity of continuous lane markings or delineators, however construction zone barricade delineators have been assessed using glance legibility techniques.

It was recommended that conspicuity be tested in situations where there are serious consequences if signs, markings or delineators are missed, and where they compete with other stimuli for attention. Static signs can be tested using slides shown for a fraction of a second and asking subjects to report features seen. For conspicuity of lane markings in snow or rain, a field test using observers to record lane encroachments was recommended.

Driver-Car Response Testing

Driver-car response measures such as speed change and lane positioning were found to be unsuccessful in testing response to static signs.

A number of techniques were used to assess driver-car response to changeable message signs (no studies were found on dynamic symbol signs). Where traffic was diverted to alternate routes because of tie-ups due to special events or construction, license numbers were recorded as cars passed the various messages and license numbers of vehicles exiting after passing the messages were recorded to establish exit volumes directly resulting from each changeable sign message.

Where traffic diverted due to unpredictable events such as accidents, traffic counters were used to estimate exit volumes. In both cases questionnaires were sent or given to drivers who had seen the message sign.

Car-driver response to lane-marking and delineator treatments has been evaluated by measuring lane-position error both in simulators and in the field. Field studies have involved the use of a car instrumented with a lane-tracker as well as the use of observers to count number of lane encroachments for traffic passing delineation treatments at construction sites.

Driver-car response testing of static signs was not recommended because it was unsuccessful and other driver response characteristics were found to be more sensitive to differences between signs. Field studies where traffic volumes exiting due to changeable message signs were recommended, the type of study depending on whether the need for diversion was predictable (e.g. special event) or unpredictable (e.g. accident). Driver-car response testing of lane markings and delineators was not recommended because it is too expensive to be practical.

General Considerations in Evaluation

A number of issues were raised that require attention when any of the above techniques is applied in evaluating a highway system. These are the following: the driver population must be adequately sampled; degraded visual conditions where an interaction is possible between signs or markings being compared and reduced visibility; a loading task to simulate the driving situation should be used, and finally, dynamic rather than static techniques should be used where possible. Whether a sign or marking is selected after evaluation will also depend on the minimum standards for comprehension, legibility distance, etc. set by the appropriate authorities.

Test Facility

In reviewing the literature it became clear that almost all the assessment methods discussed could have been administered using a combination minicomputer, software graphics and video recorder system.

Such a test facility would have many advantages. For example, it would be extremely versatile. Comprehension testing of symbol signs could be carried out using preference ranking (if at an initial stage of testing) or a multiple choice test of recognition. Legibility distance could be tested by having a sign, static or dynamic, zoom towards a

subject at a rate which simulates the speed the driver would normally approach it on the road. Conspicuity of a sign could be tested by filming the sign in its normal environment and then presenting this on a video tape recorder together with a questionnaire on the computer system. Processing time could be tested measuring glance legibility, and by having the image appear briefly on the screen. If an appropriate tracking test were developed using a simplified steering wheel, accelerator or brake pedal (as in many common video games) some of the driver-car responses could be measured e.g. erratic braking because a lane closure sign on a highway was seen too late. Driver-car response to changeable message signs could be measured by simulating a trip along a freeway allowing the subject to move at speeds appropriate to his or her choice of remaining on the highway or taking an alternate route. These examples are only a few of the ways such a test facility could be used. In the opinion of the authors all the initial testing prior to field tests should be carried out with such a test facility.

There are numerous other advantages, besides the flexibility discussed above of a computer graphics / video tape recorder system. The system could be designed to be portable so that different subject groups could be tested at different locations. Each subject could participate in several different evaluations, saving time and money involved in recruiting many subjects. Scoring of tests could be automated. Data collected on floppies could be transferred to mainframe computers for statistical analysis. Dynamic testing is possible -- signs and lane markings can be seen in motion, just as in the real world. A tracking task could be added to simulated steering and add to the realism of the test situation.

The feasibility of the test facility is mainly dependent on two features: a high resolution visual image (especially important for legibility testing) and sufficient processor speed to allow dynamic presentation of signs and land markings in real time.

Technology is certainly at a point where a test facility such as has been briefly described above is feasible. The initial costs in labour and in equipment of setting up such a facility would be offset by a reduction in the cost of evaluations and by an improvement in the highway system because in-depth assessments could be conducted quickly and efficiently.

DRIVER RESPONSE TO VISUAL HIGHWAY INFORMATION SYSTEMS – PHASE 1

CHAPTER 1 : INTRODUCTION

This report is divided into three chapters. The first chapter gives an introduction and a description of how the literature search was conducted. In the second chapter, literature on evaluation techniques is reviewed for each of the five types of driver response and each of the four highway visual information systems. In the third chapter recommendations are given for methods of evaluating each of the five aspects of driver response.

As will become clear to the reader, no one package of techniques could be recommended to cover every situation. Special considerations for evaluating each aspect of driver response are discussed. Three appendices are attached, the first is an annotated bibliography with detailed information on how a particular technique was used, the second, a list of descriptions of evaluation techniques,and the third, a discussion of statistical considerations for estimating numbers of subjects required for each evaluation procedure.

The emphasis in this report is on evaluation techniques and not on design. For this reason the conclusions of the studies reported are given only when they affect the application of the technique. For example, Gordon (1982) used miniaturized signs to study legibility distance. The results of the study showed that capital-lower case directional signs have 10% better legibility that block signs. Since the results do not affect application of the evaluation technique only the technique itself is described.

Literature Review

Literature selected for review was written mainly within the last twenty years. For literature prior to 1965 the reader is referred to 'A Study of Traffic Sign Requirements' by Forbes, Snyder and Pain published in 1964. Three sources were used to obtain literature on evaluation techniques. First, an annotated bibliography of 100 abstracts on sign evaluation written in the last 10 years obtained using the International Road Research Documentation was provided by Dr. Bob Dewar, a consultant on this project. Secondly, computerized searches were made using the DIALOG data base for literature on dynamic signs, lane markings and delineators. Thirdly, relevant literature available in the MTC library was used.

Table I shows how many references were found on each driver response characteristic and each highway information system. As can be seen, the distribution of the literature is very uneven. Static signs have received the most

attention, probably because there is such an enormous variety in use. Because so many signs are symbolic, comprehension is important; much literature was found on this topic. Dynamic signs, which include changeable message and pictogram signs and flashing arrow signs, are relatively new innovations. Consequently the research on evaluating them is at an early stage; much of it involves driver attitude towards the signs and driver response in terms of changes in traffic patterns. Very little literature was found on evaluating driver response to lane markings or delineators. This is probably due to the fact that there are fewer choices possible for lane markings or delineators than for signs, also they are continuous in nature (except for construction barricades) making evaluation of driver response characteristics such as processing rate and legibility distance of low priority.

TABLE I : Number of References

Highway Systems	Driver Response Characteristics: Legibility Distance	Comprehension	Processing Time	Conspicuity	Driver-Ca Response
Static Signs	8	9	8	5	2
Dynamic Signs	1	5	3	1	4
Lane Markings	0	3	1	0	2
Delineators	0	1	1	1	2

CHAPTER 2 : LITERATURE REVIEW OF EVALUATION TECHNIQUES

SECTION I : EVALUATION OF STATIC SIGNS

LEGIBILITY DISTANCE

Changes in Sign Size

Jacobs et al. (1975) measured threshold legibility distances for symbolic and verbal signs by manipulating their size to represent regular sized signs at distances ranging from 22 m. to 385 m. Twenty-seven different sizes were used. The effect of blurred vision on sign legibility was also examined by defocussing the lens of the apparatus showing the signs.

Gordon (1982) also used miniature signs to examine legibility distance of route numbers, cardinal direction indications and place names. He justified his use of miniature signs by citing evidence that the legibility obtained using small replicas can be extrapolated to full-sized signs. Opthalmological studies show that visual accommodation is not a problem, even the far-sighted person can accommodate as close as 1 meter (Zoelthout, 1947). Forbes et al. (1950) demonstrated constancy of threshold visual angle, that is the visual angle at which 50% of responses to the sign are correct. These data show that when letters of heights 12.7, 20.3, and 30.5 cm (5, 8, and 12 in.) are used, there is an almost linear relationship between letter height and legibility distance. If anything, the miniature gives a slightly conservative estimate of legibility distance.

Moving Subjects Toward Signs

Rather than varying the sign size, Ells et al. (1980) evaluated legibility distance by moving subjects in a wheelchair toward miniature versions of the sign to determine the relative distances at which various messages could be identified. Previous work (Dewar and Ells, 1977) had shown that the ordering of legibility distances for the signs was approximately the same using this measure as when legibility distances were obtained in an on-the-road study. The disadvantage of the wheelchair technique is that only one subject can be tested at a time and a large laboratory is required. In an early study of legibility distance, Forbes and Holmes (1939) had groups of subjects walk towards full scale signs in the field. This study used individual letters rather than words as stimuli.

Dynamic Presentation of Signs

Desrosiers (1965) conducted an experiment to validate laboratory tests in which moving picture techniques were compared with a field research method. A direct comparison was made between legibility distance measures using the two techniques. The stimuli were guide signs made to 1/3 scale of the normal size for a freeway sign. The field task was conducted by having subjects drive at 20 miles per hour (which would simulate approaching at 60 miles per hour, since the signs were 1/3 normal size) down an unused section of freeway toward the sign. The subjects' task was to indicate, by pressing a button, on which line of the sign the target name was located. The laboratory test was similar but involved a film presentation of these same signs at that location. The laboratory test and the field research showed essentially the same trends, but the mean legibility distances were five to six times greater in the field test. This large difference in legibility distance may be due to the greater illumination present in the field situation. One drawback of film simulation is that the range of luminance which occurs in real life is much reduced on film. Also, there is a problem with clarity of the image. The camera does not fixate on the sign. Instead it is directed ahead in contrast to what a driver does as he or she approaches a sign to be read.

Legibility Distance and Driver Processing Time

One result of evaluating legibility distance using a dynamic presentation is that driver information processing time becomes a factor. In a static presentation, using various sizes of signs to assess legibility distance, a sign with a long message will be legible at the same distance as a sign with a short message (for instance, one panel versus four on an overhead guide sign). However, when the presentation is dynamic, the subject takes longer to process the information on the sign with the longer message and thus

the legibility distance which is measured is shorter. Since the real-world involves dynamic presentations, this is another argument for avoiding the type of evaluation technique used by Jacobs et al. (1975) described above.

Dewar and Ells (1974b, 1979) used both reduced and full-scale signs, to compare legibility distances measured on the highway with reaction times needed to identify signs in the laboratory. They found that reaction time in the laboratory was shorter for symbolic signs and was significantly, although moderately, correlated with the legibility distances measured on the highway. The authors recommended that both legibility distance and reaction time measures should be taken whenever a complete evaluation of sign adequacy is undertaken.

Legibility Distance Under Degraded Visual Conditions

Sivak et al. (1981) examined the effect of driver age on nighttime legibility of highway signs. They used two groups of subjects, 12 people under 25 and 12 people over age 61, all of whom had normal (6/6) vision in daytime luminance. Using a field test situation where subjects were required to identify the orientation of the letter E as they drove toward it at night, it was found that legibility distances for the older subjects were 65 - 77% of those of the younger subjects. This emphasizes the fact that individual differences and degraded visual conditions are important in evaluating legibility distance. Despite this, most of the research described in this section used young subjects with good vision who viewed signs under ideal circumstances, not as they would appear in rain, snow or fog, or at night.

COMPREHENSION

More literature was found on evaluating comprehension of signs than on evaluating any other characteristic. Techniques used to measure comprehension included recognition, recall, multiple choice, matching, associative strength, preference, clarity of meaning, and semantic differential ratings. A number of studies compared different techniques mainly in an attempt to determine if recognition scores could be predicted using a technique which was more easily scored than a recognition test. Most studies examined comprehension of symbolic rather that verbal signs. However in many cases the same techniques which were applied to symbolic signs could also be used to assess verbal signs.

Recognition Tests

Dewar and Ells (1974a) compared recognition and multiple choice methods for evaluating symbol comprehension. In the recognition test subjects were asked to write out the meaning of 70 symbolic traffic signs. Percent correct responses were

scored. An answer was judged correct if the subject indicated he or she clearly understood the general meaning of the sign. While most answers were easily scored, there were occasions when some judgement was called for on the part of the examiner.

In the multiple choice method subjects were asked which of four possible meanings for each symbol was correct. A very important aspect of a multiple choice test is that the alternate choices to the right answer be reasonable ones. To obtain good alternate answers Dewar and Ells questioned 15 subjects on the symbol meanings. They also asked each subject to give additional possible meanings for each symbol to generate alternate answers on the multiple choice test. Using this method of selecting the multiple choices, the results of the multiple choice test and the comprehension test were in good agreement.

A third method tested by Dewar and Ells used a matching technique in which subjects indicated which of three symbols represented a specific message. The results of this were in good agreement with the other two methods just described. This technique examines confusability of a sign with other signs.

Another recognition technique which also looks at confusability is described by Green and Pew (1978). They evaluated 19 pictograms for use in automobiles. They measured associative strength by having subjects listen to a few sentences describing a driving scenario and then choose one of 19 pictorial symbols representing the appropriate car function (e.g. it's dark and you want to turn your headlights on). The matrix produced using this technique shows how often pictograms were confused with others and were incorrectly selected.

With each technique there are tradeoffs. The advantage of the recognition method is that, in having people write out their answers, one finds out what misconceptions people have. This can be very important. For example, in an English study a significant percentage of subjects thought that a symbolic sign meaning 'no left turn allowed' actually meant 'left turn permitted'. Depending on the consequences of the misunderstanding, a redesign of the sign might be considered. The disadvantage of the recognition method is that it is time-consuming to score and would be difficult to automate. The multiple choice method is easily scored, but requires more time to set up. The matching and associative strength methods examine the question of confusability within a system of symbols. Ideally a new sign should not be confused with signs already in use. The matching technique is more easily scored and administered than the test of associative strength.

Rating Measures

Comprehension of signs has also been measured by using ratings of various kinds. Subjects may be asked which of a number of possible versions of a sign best conveys the intended meaning. Subjects may rank order signs according to their preference or may rate the sign on a scale of 1 to 5 on clarity of meaning or communicativeness. Dewar and Ells (1974a) found that the results from the rank order and clarity of meaning procedures were sigificantly correlated.

Roberts, Lareau and Welch (1977) used several rating measures to evaluate both symbolic and verbal versions of 19 traffic sign messages. The measures were preference (rank ordering of how well the symbols conveyed a specific message), accuracy (correctness of interpretation of the sign's meaning) and certainty (how certain the subject was of the sign's meaning). None of these measures was meaningfully correlated with any other.

Using multiple choice and rating techniques, Burg and Hulbert (1960) tested 4 highway signs, each of which indicated a lane drop was about to occur. Each sign was properly installed on the highway and filmed from a moving vehicle. The filmed drives thus obtained were presented (in random order) to groups of drivers who responded, after each, to questions designed to determine how accurately the intended sign message was perceived. Signs were also rank ordered according to preference and rated according to clarity. The most comprehensible sign was the one which received the highest rank order and clarity rating. The most interesting aspect of this study was the technique for evaluating the signs. The 10 minute film strip was shown to various groups, the largest of which was 154 students. Thus a great deal of data was obtained economically and in a very short time.

More recently, Hulbert et al. (1979) and Hulbert and Fowler (1980) used this film strip technique to evaluate signs and pavement markings for the American Automobile Association. In both studies large groups of subjects in different areas of the country were assessed. A multiple choice questionnaire was administered and subjects used special marker pens with an ink which revealed immediately to the driver whether or not the correct answer had been chosen. Thus subjects were educated about the sign meaning by receiving this immediate feedback.

Semantic Differential

Determining the comprehension of a symbol by asking subjects what they think it means may be an inadequate approach when developing or evaluating a completely novel symbol. Asking the meaning of a series of very unfamiliar symbols may simply indicate that such symbols have little or

no meaning for subjects. However, it may well be that certain "unfamiliar" symbols are more "understandable" than others. A procedure which may be appropriate for such evaluation is the semantic differential, developed by Osgood, Suci and Tannenbaum (1957). This device was designed to measure the degree of 'psychological meaning' of an object or concept. Subjects rate stimuli on each of a number of bipolar adjectives. Research on the semantic differential has identified four factors -- evaluative (e.g. good-bad), potency (e.g. strong-weak), activity (e.g. active-passive) and understandability (e.g. strange-familiar)-- associated with psychological meaning. Since comprehension of symbols may involve such dimensions of meaning as understandability and evaluation, the possibility that this measure of meaning might reflect symbol comprehension was explored by Dewar and Ells (1977). This test was administered using a variety of symbols, some of which were familiar and some of which were unfamiliar to Canadian drivers. Prior to having subjects rate each symbol on the semantic differential, comprehension of the meaning of each was determined by having subjects write out what it meant (recognition method). Semantic differential scores on all four factors were significantly correlated with degree of understandability of the symbolic messages. It was concluded that the semantic differential can be readily used to gauge comprehension of novel or unfamiliar symbols, and may be a useful tool for the development of new symbolic messages. A possible advantage of this method is that more thought is put into the responses made in the semantic differential test, as compared with other rating scales, since the subject rates the stimulus on at least three adjective pairs corresponding to each semantic differential factor. This could provide a more reliable measure of the meaning which the stimulus had for the subject. Information about the psychological meaning of these apparently meaningless stimuli would be useful in eliminating certain members of a set of possible symbols which had been designed to convey a specific message. As used in the past this method is very time consuming, however it could be speeded up by computerization.

While this method appears promising, it has only been tested with highway signs by one researcher on one occasion. Further research would be necessary before it could be used as an evaluation tool.

Use of Multiple Measures for Testing Comprehension

A number of studies have addressed themselves to the issue of the best and most efficient combination of tests for measuring comprehension. Easterby and Zwaga (1976), in a study of international symbols, recommended the rank order preference technique to select three versions of each of several messages for further testing by the recognition and matching procedures. The version to be finally adopted is decided by the outcome of the comprehension test.

Green and Pew (1978), in a study of pictograms for use in automobiles, used three comprehension tests, familiarity, associative strength and ratings of communicativeness as well as a paired-associate task to measure learnability and a reaction time task to measure processing rate. These authors recommended the rating of communicativeness since this measure was best correlated with associative strength and reaction time measures.

Cairney (1982) discusses the Australian standard requirement for symbol testing in light of a series of experiments and makes a recommendation similar to Easterby and Zwaga's. The Australian standard requires the preference ranking test to select one version of the sign which is then subjected to further testing for comprehensibility and ease of learning (recall test). It is on the second of these criteria that the decision whether or not to adopt a particular symbolic sign is taken.

Cairney criticized this selection procedure on the grounds that by eliminating all but one version of the sign before comprehension or recall testing there is no guarantee that the sign which is most readily understood or learned is adopted. Also, in a series of experiments, Cairney found that scores on the recall test were generally so high that they did not differentiate between signs even though one sign might be more comprehensible than another.

Although Dewar and Ells (1974a) showed good correlation between the results of the recognition test and the preference ranking procedure, this does not mean the recognition test can be dropped. Cairney points out that the proportion of correct predictions is what we are most interested in. While the correlation coefficient between different tests is widely used, it can give a misleading impression that if the results of two tests are similar enough, one test can be dropped. Cairney found that the ranking test (where subjects rank order versions according to their preference) correctly predicted the best understood sign in 11 out of 17 signs tested. For 3 signs, scores in one or other of the tests were such that no best sign could be selected. In 3 cases the ranking test predicted the wrong sign as being the best understood. In examining the rating

test (sign communicativeness rated on a scale of 1 to 5), Cairney found it to predict the most comprehensible sign in 9 out of 17 cases. In the 5 cases where the highest rated sign was not the best understood sign, the worst of the three alternatives was selected. Cairney reanalyzed data reported by Roberts, Lareau and Welch (1977). These authors tested recognition threshold, reaction time and comprehension for each of 4 all-symbolic versions of 17 signs. They also obtained confidence ratings for the accuracy of the meanings and preference rankings for the alternatives. An efficiency index for each version of each sign was compiled by summing the Z-scores of that particular version of the five variables. Cairney found than in 11 out of 17 cases the preference ranking score would have selected the best of best equal version on the comprehension test; the 'efficiency index' predicted the best or best equal version in 16 out of 17 cases.

In conclusion, though preference rating and ranking may be easier to administer than tests of recognition, their results do not always correctly predict which sign will be best understood. Studies which examine a number of tests of comprehension indicate that more than one method is necessary to completely evaluate comprehension.

PROCESSING TIME

Processing time is a measure of how long it takes a driver to comprehend a sign message. A traffic sign message may be understandable when it is seen clearly for a few seconds. However, drivers are frequently in situations in which they have much less time to understand and respond to a traffic sign. Techniques for assessing processing time include laboratory measures of reaction time to verbalize a response, percent of correct responses to a sign shown for a fraction of a second (glance legibility), and field measures using eye-marker cameras to measure how long a driver actually fixates on a sign to read it.

Ells and Dewar (1979) discussed the evolution of their technique for measuring reaction time in the laboratory. In earlier studies (Dewar and Ells, 1974b), reaction time was measured between the appearance of a sign and the beginning of the subject's verbal response. However it was felt that responding to a symbolic sign was taking longer than it should simply because translation of the symbol meaning into words was required. To improve the validity of the technique a 'same-different' reaction time paradigm, rather than verbal identification, was used. This procedure required subjects to respond aloud with the word "yes" or "no", depending on whether a projected sign message was the same as or different from the orally presented referent sign message. The referent message preceded the traffic sign slide by approximately one second. Response errors were kept to about 5% by emphasizing

to the subjects the need for accuracy. The mean reaction times were significantly correlated with the previously determined field legibility distance data from a roadway study for both verbal and symbolic signs.

Glance legibility is another common technique for measuring processing time. This technique measures number of correct responses to a sign presented for a fraction of a second. Such a measure would seem appropriate, since drivers often have time for only a brief look at a traffic sign, especially while driving in heavy traffic on urban streets. Dewar and Swanson (1972) used the glance legibility technique described above and showed symbolic signs are more easily identified than word signs containing the same messages.

A variation of Dewar and Swanson's technique was used by Markowitz et al. (1968). In order to provide a measure of recognizability that was uncontaminated by subject biases or predispositions, subjects viewed signs presented for a fraction of a second and then identified, from a set of candidate signs, which one had been presented. Dietrich and Markowitz (1972) later used this method to evaluate the effectiveness of a number of new signs which had recently been introduced into the United States system.

The most sophisticated and expensive technique for measuring information processing employs eye-marker cameras. Bhise and Rockwell (1973) describe the development of a methodology using an eye-marker camera for evaluating road signs. The eye-marker camera system provides continuous records of the driver's eye movements (i.e. where the driver's eyes are directed while driving) superimposed on the driver's view of the forward road scene. Data were collected on the eye movements of drivers under actual driving situations involving more than 400 different highway signs. The data were used to evaluate signing characteristics such as letter size, length of message, relevance of message with respect to exiting or route-following instructions etc., and driver characteristics such as familiarity with the road, visual acuity and workload due to various levels of traffic density.

Measures were made of the maximum time available for the driver to read the sign, i.e. the time during which the sign could form a resolvable image on the driver's retina if the driver were fixating directly on the sign, as well as the time actually used by the driver to read the sign. Two of the most important findings of this research were that first, drivers did not just concentrate on a sign (i.e. read a sign in one glance) but rather made several glances at it. Secondly, the time sharing process of the drivers with the signs and other objects on the road was found to be dependent on factors such as time-distance to first fixation on the sign, traffic density, type of information needed by the driver and driver familiarity with the road.

The eye movement evaluation technique described above does take account of a very important aspect of driving which is ignored in most studies of sign perception, namely that the driver must time share between reading signs and driving. This time sharing requirement is particularly important in evaluating processing time. In busy traffic a driver will require more total time to process a sign than in light traffic because more time is taken up watching for and responding to other cars.

Noble and Sanders (1980) used a compensatory tracking task in conjunction with a visual search task where response time to identify familiar signs out of a group of signs was measured. Both the search task and the tracking task showed degraded performance when they were performed concurrently.

Testin and Dewar (1981) examined reaction times to signs both while the subject was performing a loading task and when the subject responded only to the signs. Reaction time to the signs increased in the presence of the loading task. However the correlation between reaction time and legibility distance (measured in a previous experiment on the roadway) did not improve in the presence of the loading task. This may be due to the fact that the loading task used was initiated just prior to the sign reading task and it was discrete rather than continuous in nature. A tracking task would more closely simulate the demand of the steering task on the driver's attention and would be a better loading task in the laboratory. It remains for future researchers to evaluate reaction time, using a more appropriate loading task in the laboratory, as a predictor of legibility distance in the field.

Processing Rate Under Degraded Visual Conditions

The effect of degrading sign visibility in the laboratory was also measured by Dewar and Ells (1977). In the non-degraded condition, subjects saw the slides through goggles with transparent noncorrective uncoloured glass lenses. In the degraded condition, the same slides were viewed through the lenses of the goggles plus a lens cover composed of layers of plastic film. Light bulbs on either side of the subject's head, were illuminated during both conditions. In the degraded condition, the lights in conjunction with the plastic lens covers generated a bright veiling glare between the subject's eyes and the projected sign. The glare was similar to that produced by automobile headlights in fog and darkness. A result of this experiment, which shows the importance of including the degraded visual situation, is that reaction time to verbal signs was much more affected by visual degradation than was reaction time to symbolic signs.

CONSPICUITY

Conspicuity is of great practical importance, but its measurement presents some difficulties. The problem is that it has been defined in different ways and measurement techniques reflect different operational definitions. More than any other sign characteristic, conspicuity focuses on the context or background in which the sign is placed.

Forbes, Fry, Joyce and Pain (1968) defined conspicuity as attention value, two components of which are target value -- the extent to which a sign stands out from its background and priority value -- the attributes which determine the order of noticing signs of equal target value. In this study a set of signs were presented for 1 second while the subjects performed a secondary task. The subjects were asked which sign they saw 'first and best'.

Cole and Jenkins (1982) used the following operational definition of conspicuity: 'a conspicuous target object is one that will attract attention and will be seen with certainty within a short observation time, regardless of the location of the target object in the visual field'. This definition resulted in three experimental techniques. In all experiments, subjects were shown colour slides depicting various signs in a variety of daytime urban or semi-urban road scenes. Each slide included an obvious target sign.

In the first experiment, slides were projected for half a second and the subjects' initial fixation point was fixed. Subjects were asked to report the road fixtures they saw that were pertinent to them as drivers. Subjects wrote down each feature seen, its location and the message displayed or other pertinent details. The response enabled a distinction to be made in later analysis between 'detection' and 'recognition'. The half second exposure time allowed for one exploratory eye movement. The second experiment was identical to the first except that a longer exposure time of one and a half seconds was used to allow more features to be recognized. In the third experiment, four observers experienced in visual psychophysical experiments were shown each slide for as long as necessary and were asked to judge the primary target feature and its background with respect to 9 attributes: target and background, brightness and colourfulness, target size and boldness, target edge contrast and sharpness, and background clutter. Although some of these physical attributes could be measured, the authors felt that the relation between physical quantity and perceptual response are not well understood, especially for complex visual environments.

Some of Cole and Jenkins' findings emphasize the need for conspicuity testing. They found that none of the primary target features could be considered to be highly conspicuous since none served to attract first attention of all subjects

on all occasions. Small signs with white backgrounds were found to be particularly inconspicuous. Increasing the observation time in the second experiment by 300% only resulted in a 35% increase in number of features noticed. A strong correlation was found between estimates of conspicuity at half a second and of probability of detection at one and a half seconds, a more realistic viewing time.

Woltman, Stanton and Stearns (1984) tested conspicuity of highway guide signs using a video presentation. Their experiment concerned the conspicuity of white-on-green versus white-on-black signs in nighttime settings. Six nighttimes scenes of various background complexities were photographed. Miniature guide signs were photographed by a second video camera; background and sign images were then superimposed to form a composite image. Scenes were constructed with all white-on-green or white-on-black signs. A total of 313 subjects then viewed the scenes, presented by a video recorder, for 3.5 seconds and counted the number of signs. The technique allowed different signs to be compared in identical backgrounds.

Field studies of sign conspicuity have involved having subjects report each traffic sign they pass (e.g. Summala and Naatanen, 1974). Very high reporting rates are found. However eye-movements studies (Mourant, 1969) have shown that drivers' eye movements are very different from normal when they are required to report all signs. Another technique used in field studies has been to stop drivers at the roadside and ask them to recall the last sign they passed (Johansson and Backlund, 1970). In contrast to the results of the first type of study, very low rates of recalling signs were found. Signs which were recalled best were those that were of greater importance to the driver, indicating that motivation will have some effect on conspicuity and on the attention value of a sign. There are major drawbacks to this technique. The driver may be so surprised at being flagged down that he or she forgets the sign just passed. This does not mean that the driver did not see it. Also responses of drivers familiar with the area are likely to be different those of visitors.

Each of the techniques reported above is relevant to different situations in which measures of conspicuity are required. Cole and Jenkins' (1982) technique of using a tachistoscopic presentation is appropriate in assessing signs in complex backgrounds e.g. where guide signs compete with advertising signs on a highway, or where signs need to be conspicuous in an urban setting where other signs e.g. street names, regulatory, warning or advertising signs are present. In both these situations the sign must be noticed quickly, on the highway because of high speeds and in an urban setting because of the greater number of other visual stimuli. In simpler visual conditions, the technique used by Woltman et al. (1984) of combining backgrounds and signs in a

superimposed image and having subjects count signs would be more appropriate.

DRIVER-CAR RESPONSE

Measures of driver-car response, as opposed to driver alone response, require either the use of a driving simulator or a field study. Few studies were found of driver-car response to static signs. Perhaps the fact that only a few of the static signs a driver sees require an overt response has resulted in a lack of interest in this measure.

Hulbert and Beers (1966) used a film-based driving simulator to study the relative effectiveness of various types of signs in preventing drivers from entering freeways via off-ramps. The camera car drove onto a freeway ramp the wrong way and the subjects responded by deccelerating and braking when they realized the situation. Decceleration time and distance and braking time and distance were recorded. These measures did not clearly differentiate between signs because some drivers slowed but continued on, being unsure of the meaning of the signs.

Seymour, Deen and Havens (1972),in a field study, made observations at lane closures on interstate highways to compare the effectiveness of yellow and orange signs in warning drivers of the lane closure. Traffic conflicts and merging manoeuvres were observed. Traffic conflicts were defined as abnormally violent brake applications, situations where a driver cut in to merge, and complete stops where a driver waited too long to merge and had to stop and wait for a gap. The point of merging was observed and percentage of drivers merging near the traffic cones (considered undesirable) was determined. Differences in car-driver responses between the two signs were small, however driver preference polls supported the orange signs more strongly.

In neither of the studies described above were driver-response measures successful in clearly differenting between signs. In the real-world, and in simulators, there is much variability between drivers. Thus many observations are required to detect differences in behavior. It may be that for static signs, more closely defined evaluation techniques (e.g. measures of legibility distance, processing rate etc.) will yield information about differences between signs more reliably and with fewer data than driver-car response measures.

SECTION II: EVALUATION OF DYNAMIC SIGNS

INTRODUCTION

Changeable message signs and flashing arrow signs are both fairly recent innovations on the highway. Changeable message signs have been used mainly to give incident information, that is, information about accidents or construction. More recently they have been used to give alternate route information and to replace static speed signs under conditions of reduced visibility. Flashing arrow signs are used to guide drivers through construction zones.

Because of their recent introduction, the research on evaluating dynamic signs is at an early stage. A number of studies will be described which report the results of questionnaires on driver attitudes about the type of information they would like to see on changeable message signs. There are also studies about the format the drivers prefer the message to be presented in, about the legibility of different messages, different sign types and degraded equipment (bulbs missing). There were no studies found which dealt directly with how long it took drivers to read a particular message i.e. driver processing rate. To simplify reporting of the literature, studies of driver preference for a particular sign message, in terms of how well it conveys the idea to be communicated, will be included in the section on driver attitude below. Very few literature references dealt with evaluating dynamic symbolic signs, e.g.flashing arrows.

DRIVER ATTITUDE

There were three aspects of driver attitude evaluated in the literature: 1) driver attitudes about their potential use of a changeable message sign, 2) driver attitudes about a changeable message sign they saw in operation, 3) driver preference for one sign message over another.

Driver Attitude About Potential Use of Changeable Message Signs

Three early studies reported in 1970 and 1971 looked at the probable reactions of drivers to changeable message signs. Heatherington, Worrall and Hoff (1970) used a questionnaire to solicit the preferences of 732 Illinois drivers for different traffic information messages they might be presented with. Respondents were asked to select which of a pair of messages they preferred for describing each of three levels of congestion. A total of 19 traffic messages were evaluated. For heavy congestion, information about an

accident having occurred and causing heavy congestion was most preferred. A description of expected speed ranked second for the level of heavy congestion, and first for the levels of moderate and light congestion. Delay and travel time descriptions were not desired by the respondents.

Dudek, Messer and Jones (1971), using a similar evaluation technique, surveyed 505 licensed drivers from businesses in Houston and Dallas, Texas. Subjects completed a multiple choice questionnaire. Results of the study showed motorists did desire additional traffic information not provided by static signs, that they preferred to receive the information before entering the freeway so they could divert to an alternate route, that they preferred information about the location and length of the congested area and the degree of congestion, and that quantitative descriptors of travel time and travel speed were least preferred. This last finding is in agreement with answers to the Heatherington et al.(1970) questionnaire given to Illinois drivers.

Hulbert and Beers (1971) used three simulation techniques to evaluate the type of messages drivers prefer as they approach traffic congestion. The first method involved having groups of people view film strips; in the second, a tabletop simulator was used by which means the subject viewed freeway drawings showing sign messages, and the third method used a driving simulator. The techniques sound promising but unfortunately they are not described in any detail. Driver's feelings about the general concepts of changing messages and their reaction to more than 80 specific messages were assessed.

Huchingson and Dudek (1979) conducted five studies of motorists' opinions using a pencil and paper technique. A traffic situation was described; the motorist was then shown a card describing a traffic incident followed by a card with a message such as would be displayed by a changeable traffic sign. The effects of different types of incidents (accident, roadwork, etc.) and amounts of delay on whether or not a driver said he or she would divert to another route were investigated. What drivers understood by a delay message was investigated using a scale rating of 1 to 5 to indicate agreement with each of five interpretations of a traffic sign message. Major findings were that 50% of drivers stated they would divert from a freeway if the delay duration was at least 15-20 minutes, and would divert to a bypass route if the displayed time saving was at least 5-10 minutes.

Driver Attitude About Signs in Use

Three evaluations of changeable message signs combined a field study with a questionnaire to examine driver preferences. In 1976 the Texas Transportation Institute conducted a series of field studies of point-diversion of freeway traffic going to special events in Dallas. License

numbers of cars observed during the study were recorded. Two questionnaires were sent out one month after the field study and the return rate was 37%. Since the signs the drivers were asked about were novel and the messages related to a specific event, a delay of one month may not be a problem. In other circumstances, such a long delay might have resulted in too low a return rate. Richards, Stockton, and Dudek (1978) describe the development of the two questionnaires, one sent to drivers who diverted in the field study, the other sent to drivers who did not divert. The questionnaires attempted to obtain driver attitudes and reactions to the point - diversion signing systems, explored trends in driver behavior related to the diversion decision, and identified factors that influenced a driver's decision to divert. One of the findings was that the advisories of severe traffic problems were cited most often as the information that would have influenced those drivers who did not divert to do so. This finding was suprising because some of the drivers did in fact see messages of this type but must have failed to read them or misinterpreted them. However, the finding that drivers would have diverted if they were told severe traffic problems lay ahead is in agreement with questionnaire only studies by Dudek et al. (1971) and Heatherington et al. (1970).

Turner, Dudek and Carvell (1978) report another combined field and questionnaire study. In the field study freeway traffic was diverted during maintenance operations; also, a nine-camera closed circuit television system was used to give information about traffic accidents. The effectiveness of the various messages in diverting traffic were measured by: 1) the change in diversion rates from natural diversion (no sign) to diversion because of information on changeable message signs and 2) questionnaires, which were distributed to drivers who actually passed the sign displays. How this distribution of questionnaires was actually carried out was not reported, nor was it reported whether questionnaires were filled out on the spot or mailed back. Drivers were asked to evaluate the information displayed as well as give their opinions about what further information would be helpful. One interesting finding was that, in the field studies, exit volumes were consistently greater for diversionary signs (take alternate route) than for informational signs (congestion ahead), which would seem to indicate motorists acted on diversionary information. However, the questionnaire results indicated that none of the motorists who received alternate routing information cared for this type of information. Thus there was some disagreement between field study results and questionnaire results. Either driver opinion doesn't always reflect driver behaviour or else that drivers who took an alternate route, on the advice of the sign, were not happy with the route suggested. Therefore, one must be careful about drawing conclusions based solely on driver opinion.

A third report of a combined field and questionnaire study was made by Hanscom (1981). Right and left lane closure situations were studied and observations were made under day and night conditions in four states in the United States and with about 200,000 vehicles. The 489 drivers who answered the questionnaire in this study were actually recruited as subjects in the field experiment. This is in contrast to the studies reported above where drivers were not specifically recruited for the study. Substantial proportions of drivers younger than 20 and older than 60 were recruited as well as nearly even numbers of male and female subjects in an attempt to approximate the normal driving population. Driver questionnaires containing 20 questions were administered immediately after the subject finished driving through the study area. The questionnaire strategy involved asking a series of general questions regarding observations of traffic control devices which the drivers had passed. Although answers to these questions were provided prior to the subjects being directly asked about their changeable message sign observations, the answers nevertheless reflected an indirect impact of the signs, and thus afforded an internal response validation for the questionnaire.

COMPREHENSION

In a study of flashing arrow boards, Knapp and Pain (1979) point out that most of the literature deals with descriptions of different types of arrow boards and studies showing that arrow boards are effective in diverting traffic and are highly visible. However, Knapp and Pain argue that because of the varying arrow configurations and sign placements, evaluation is needed of which combination of these is most effective. Two studies were conducted. In the first experiment subjects were shown 9 film clips, each one representing different arrow boards placed on the lane or on the shoulder. After each film clip subjects chose 1 of 4 responses indicating the meaning of the arrow and also gave a rating of his or her confidence in the response. The second study was a refinement of the first. More subjects were used (109) and the film clips were improved so that the difference between shoulder placement and lane placement was clearly seen. Six short film clips were shown, presenting two arrow configurations simultaneously, side by side. All possible pairs (6) were evaluated and subjects chose the preferred configurations in each case.

At the Transport and Road Research Laboratory in England researchers have developed 'pictogram' messages for various roadway hazards ahead and presented them using different types of changeable message signs. Rutley and Webb (1979) describe laboratory investigations on recognition and meaning of these displays. Groups of subjects viewed the 'pictogram' messages at a test track from the side windows of a bus

parked within easy viewing distance. Each symbol was displayed for 7 seconds and subjects wrote down their interpretation of each signal. Interpretations were classified as correct, similar meaning, wrong or no answer. A second trial using a 30 second presentation time did not find any significant changes in level of understanding. The pictograms similar to signs already in use were best understood, new pictograms were understood by 24 to 91 % of the subjects.

Mounce, Messer, Huchingson and Dudek (1977) evaluated bulb-loss effects on message readability. Ninety-three subjects gave written responses to presentations of matrix signs with varying amounts of bulb-loss. A representative group of traffic-description and advisory words and route numerals were displayed on a real time matrix sign. Selected percentages of bulbs were failed in a random fashion and slides were taken of the resulting displays. The slides were shown to 93 subjects of varying ages, education and driving experience, representative of the normal driving population. Subjects were instructed to respond by writing the word if it was legible. From these data, specifications for 85th and 95th percentile correct comprehension were determined for motorists familiar and unfamiliar with the presented displays.

With the exception of the above reported study on bulb-loss effects, the only literature found where comprehension of verbal changeable message signs was measured used preference ratings; these studies are described earlier in this chapter in the section on driver attitude. Recognition, that is, drivers' understanding of the meaning of verbal signs, does not appear to have been considered in evaluating changeable message signs. Perhaps the assumption has been made that the drivers have complete understanding of the signs, abbreviations included, which have been presented so far. The only study to even touch on this question was by Huchingson and Dudek (1979) who asked drivers to estimate how long a delay they thought the messages 'minor accident' and 'major accident' would indicate. Considering that materials in two studies cited earlier indicated drivers did not like messages giving actual delay times, it would be worthwhile ensuring that delay times associated by sign designers with specific messages are also those associated with such messages by drivers. Also, where diversion information is being presented to drivers it would be worth evaluating whether drivers understood which alternate route was being indicated and understood any abbreviations that were used.

LEGIBILITY DISTANCE

Three studies, one of letters in a dot matrix display in the laboratory and two of pictograms in the field, were found in which legibility distance of dynamic signs was

evaluated. Snyder and Taylor (1979) examined various techniques for evaluating the legibility of alphanumeric characters which were presented in a dot matrix display. Though changeable message signs per se were not evaluated, the techniques discussed would certainly be appropriate for such an evaluation. Snyder and Taylor used a CRT to present alphanumeric characters to subjects. Display parameters investigated included character size, luminance level and viewing distance. A computer program was used to generate random presentations of 18 alphanumeric characters under each experimental condition (character size, luminance level and viewing distance). Subjects were trained on the tasks prior to data collection. At each session the subject sat in a dark room in front of a CRT at one of a total of seven viewing distances. When ready to begin the subject pressed a hand-held button. The terminal was used to display instructions to the subject, asking him or her to use the button to initiate each trial and to respond as soon as he or she recognized the alphanumeric character presented. The subject then pressed the console key of the alphanumeric character presented. Where viewing distances were large, and the subject was far from the console, an experimenter pressed the console button according to the subject's instructions.

By these means, measures of accuracy, response time and threshold visibility (viewing distance at which 85% of the characters displayed were accurately recognized) were made. Although this research was on single alphanumeric characters, the technique readily lends itself to evaluation of changeable message signs. One advantage of the use of the computer to conduct these tests was the reduction in workload for the experimenter and the quick and accurate scoring of responses. Although only six subjects participated in this experiment, these factors would allow a larger number of subjects to be tested than is possible in non-computerized laboratory set-ups.

Rutley and Webb (1979) used a prototype signal to measure legibility distances of pictograms in a field set. The effect of legibility distance of a matrix signal as it becomes dirty was also assessed during this trial. Plain, perspex screens had been placed for six months on a highway, in roughly the same position as signals would be, to accumulate dirt. These were placed in front of the test pictograms for half of the trials. Before the field test subjects were shown photographs of the pictograms and their meanings were explained. The subjects (24) were then driven individually toward the signal on the test track. As the subject was driven slowly toward the signal he was asked to identify the legend as soon as possible.

On the first viewing some of the pictograms could only be recognized at short distances; on the second viewing recognition distance increased by an average of 33% emphasizing the need for some practice for the subjects

before recording results. In previous work (Hodge and Rutley, 1978) bright sunlight prevailed and recognition distances for numerals were 25% shorter than in this study where skies were cloudy. The dirty screens caused no statistically significant reduction in recognition distances even though overall light was reduced by 32%.

While the studies reported above considered the effects of dirt and sun versus cloud, they did not evaluate signs under other conditions of visual degradation such as rain, snow or fog. This should be considered in future evaluations because it is often under these conditions that a changeable message sign will be most valuable to the motorist.

CONSPICUITY

The only evaluation of conspicuity found was the simple technique used by Hanscom (1981). Almost five hundred drivers were recruited to drive past test sites with changeable message sign installations showing various messages. In a questionnaire, drivers were asked whether or not they saw the message that was being evaluated. Percentages of drivers reporting in the affirmative ranged from 67 to 87% where one sign was used, and from 78 to 93% where two signs were used. While this indicates some messages are more conspicuous than others, no reasons for this difference were determined.

DRIVER-CAR RESPONSE

Changeable message signs have mainly been used to divert traffic around construction zones as well as to alternate routes when the main route is overloaded because of congestion due to an accident or some special event. Because of this much of the research on them has focused on driver-car response. Much work in this area has been carried out by the Texas Transportation Institute and has involved observation of traffic speeds and volume of traffic diverting off freeways.

Stockton, Dudek, Fambro and Messer (1975) reported a study using changeable message signs to inform motorists of freeway traffic conditions. A letter-grade rating scale for indicating freeway conditions was also evaluated. Both traffic flow observations and questionnaire surveys were used to evaluate the signs. Freeway volumes and speeds, and ramp volumes were input from strategically placed detectors to a computer in the surveillance centre. These data were recorded each time a 'lane blocked', 'slow traffic' or 'OK' message was displayed on the freeway sign.

To assess the effect of voluntary diversion (i.e. that due to evident congestion), data were obtained after the

signs were removed, during periods when the 'lane blocked' sign would have normally been on. These two sets of data were then compared to ascertain the portion of the traffic diverting as a result of information displayed on the signs. Questionnaires were also mailed to drivers observed passing the freeway sign during certain message displays. Questionnaire data formed a useful addition to the traffic oservations. (See Driver Attitude Section).

In 1976 and 1977 Dallas freeway traffic for various special events was diverted to alternate arterial routes using changeable message signs. The evaluation technique reported by Weaver, Dudek, Hatcher and Stockton (1977) was again a combination of traffic observation and questionnaire surveys. Messages were displayed for 15 minute intervals separated by 10 minute blank conditions to allow the traffic to stabilize. The traffic situation was continuously monitored throughout the study area by using radio communication. In this way real-time information was provided about expected delays and traffic movement so that messages displayed were realistic.

To ensure that all vehicles passing the matrix sign were included in the analysis time interval, a floating car left the sign at the instant a message was activated or terminated and traveled along the freeway at the speed of the traffic stream. The time the floating car exited from the freeway to use the designated alternate route was recorded. By knowing the travel time of this car, vehicles could then be positively identified as having passed a particular message.

Comprehensive origin-destination data were collected by recording the license plates of almost every vehicle passing selected points on the main and alternate routes. License plates were read through high-powered binoculars mounted on tripods. Observers rotated every 10 minutes and provided better than a 95% count when compared to the main-line volume data from the permanent detectors. Observers also recorded the time to ensure proper co-ordination with the sign messages. License numbers were used to obtain driver addresses and questionnaires were mailed out.

The evaluation technique described above is extremely thorough and well suited to predictable occurrances such as maintenance activities and special events. The next study describes an evaluation of the use of changeable message signs to alert drivers to unpredictable incidents causing traffic tie-ups. As Dudek, Stockton, and Hatcher (1982) report, the 'random nature of incidents precludes keeping a license plate data collection crew on standby'. In this study a network of traffic volume counters was used to obtain evaluation information. Traffic counters were installed upstream from the diversion point, on three interchange ramps. Traffic counters recorded volumes at 5 minute intervals.

Analyzing these data presented some problems: changes in volumes on some of the ramps were small and it was difficult to tell if the changes resulted from the signs or from random variations in traffic destinations; also the traffic counters did not cover all possible input and output points. 'Natural' diversion had to be estimated. Previous studies (Stockton et al., 1975) have shown that a significant number of drivers divert from a freeway upstream from their intended exit points whenever unusual congestion occurs, even though they don't know the cause of the problem. Because this type of diversion can be quite high, it had to be considered in order to more accurately evaluate the true effects of the message signs. Thus Dudek et al. (1982) compared volumes for incidents where messages were used, to volumes for incidents occurring at approximately the same time of day, for which messages were not used. Volume counts were also obtained during normal (non-incident) traffic for several days (same day of the week as the incident day) within two or three weeks of the incident day. A case study incident was obtained where data were collected for all three of these situations. To account for normal traffic variations, a direct comparison of volume changes on the diversion ramps was not used. Instead, volumes on the ramps were converted into percentages of the freeway demand volumes. The basic assumption here is that the percentage of drivers who would normally use a particular off ramp at a particular time is the same for each non-incident day. Increases in traffic on the diversion ramp would be attributed to the incident and the changeable message signs.

In the studies reported above, large numbers of drivers were unknowingly observed. Hanscom (1981) reports a variation on this technique to study the use of changeable message signs for lane closures. Driver behaviour was unobtrusively observed both for drivers who were recruited to participate in the study and for drivers who unknowingly drove past the test signs. Drivers were recruited so that their behaviour could be matched to results of questionnaires they filled out (see Driver Attitude). Observations were made of traffic volume and speeds both before and after the messages indicating lane closures were in place.

Traffic volumes were sampled during morning and evening rush hours, off peak times and in darkness. Observers were stationed at several points along the test site: one mile in advance of the lane closure (before the sign was visible), at the sign, midway between the sign and the taper and 100 feet in advance of the first taper channelizing device. This last collection point was important for collecting data on how many hazardous late exits occurred. Observers alternated between sampling vehicle speeds and lane distribution (vehicle counts per traffic lane). Vehicles other than cars, such as trucks and buses, were counted separately. The recruited drivers behaved similarly to the drivers who were

unknowingly observed. The main advantage of recruiting drivers was that they all answered a questionnaire and this provided a validation for questionnaire responses mailed in by only a portion of the non-recruited drivers.

SECTION III EVALUATION OF LANE MARKINGS

INTRODUCTION

Lane markings have received considerably less attention than signs. One reason for this is the restricted range of lane markings in use compared to the enormous variety of signs. Literature found on lane marking evaluation fell mainly into two categories: comprehension and car-driver response. One study examined both processing time and confidence ratings; it will be included in the section on comprehension below. No studies were found on conspicuity of lane markings.

LEGIBILITY

Legibility of lane markings seems to have been primarily assessed by making physical measurements of contrast. However, as was pointed out by Cole, an optometrist involved in traffic research, 'the relations between physical quantity and perceptual response are not well understood, especially for complex visual environments' (Cole and Jenkins, 1982). In a field study on evaluating lane markings, Allen , O'Hanlon and McRuer (1977) had planned on using photometric measurements. However, because of time problems, visible range (legibility distance) was estimated by the experimenter who counted successive visible lane-line segments and multiplyed their number by the lane-line repetition cycle. Of course this is a sample of one and no comparisons were made with photometric measures. As a technique it is of interest, and was the only report found of a visibility measure using a human observer.

COMPREHENSION AND PROCESSING RATE

Gordon (1976) reported a study designed to examine drivers' reactions to lane markings, in particular their understanding of the road marking code in use, their opinion on which traffic situations required special road markings (e.g. approaching a narrow bridge), and their opinion on what would be appropriate markings for a variety of common highway situations.

Drivers' understanding of the road marking code in use was assessed using both a questionnaire alone and a film presentation combined with a questionnaire. The questionnaire tested recognition of road markings by showing highway scenes and asking the drivers to explain the meaning of the road marking. One of the difficulties in testing recognition is scoring responses. In this study a comprehensive code was constructed which contained all

categories of right and wrong answers. A right answer was defined as one containing no incorrect statement even though the explanation was not necessarily complete. A non-defining answer such as 'to separate lanes' was considered wrong.

An interesting finding was that lane markings were not well understood; for example, over half of the subjects tested could not correctly define the meaning of the single white line. The film presentation showed highway situations leading up to a driver decision point. After each situation the subject answered questions about permitted behavior with yes, no or don't know responses, and in some cases with responses about direction of permitted passing behavior.

The other two evaluations made in this study used questionnaires to examine driver attitude. In the first questionnaire, subjects saw pictures of traffic situations with a written explanation and were asked if special road markings were required. In the second questionnaire subjects saw diagrams showing lanes and permitted direction of travel representing different types of highways. Subjects choose an appropriate marking from a set of markings provided. Although it was difficult to make exact comparisons between the film and the recognition test, there appear to be more correct responses to the film questionnaire, indicating drivers knew what to do in some situations where they could not express the right answer verbally.

Allen et al. (1977) evaluated driver attitude to different lane markings with a rating scale to measure how demanding of the driver's attention lane keeping was, using different delineation treatments. Drivers were given the rating scale after each of three driving tasks, in a graphics-display based simulator. The scale contains descriptive phrases of driver workload, e.g. mildly demanding; excessive control demands; cannot maintain adequate control. The words and phrases were developed so that an interval rating scale would result. An interesting finding was that the drivers' ratings showed that discomfort occurred at a visibility level somewhat above the point where car control (lane exceedances) had apparently deteriorated. Thus car-driver response in this case was a less sensitive measure of performance than driver opinion. One major drawback of this study was the use of only six subjects.

Triggs, Harris and Fildes (1979) measured driver attitude about delineation treatments of curves. Slides showing curves at night with various delineation treatments were presented and subjects made a judgement as to the direction of the road, and their confidence (on a scale of 1 to 5) of their judgement. Confidence levels were significantly higher when edgelining and/or delineators were present. Reaction time to make judgements was also measured. Confidence levels and reaction time measures were correlated for road-side post delineators but not for edgelines.

The questionnaires and film presentation described by Gordon (1976) relate to understanding the road markings code whereas the papers by Triggs et al. (1979) and Allen et al. (1977) relate to comprehension of selected road markings under degraded visibility conditions. The combination film presentation and questionnaire described by Gordon seemed to have a slight advantage over the recognition test in evaluating driver's comprehension. The questionnaire also has the advantage of being more easily scored. Both methods could be and, in fact, were used with a large number of subjects. The use of driver ratings in combination with simulation by Allen et al. (1977) seemed a very sensitive measure of how degraded visibility affected driving performance. However this technique might be applied to more subjects more inexpensively if a field test situation rather than a simulator was used, and drivers were questioned at the road side. The technique reported by Triggs et al. (1979) was acceptable for the situation in which it was used, i.e. judgements of the direction of curves. Road markings tend to be continuous presentations whereas signs are discretely presented. Thus, there would likely be many instances in which a static evaluation (slides) of road markings would be less desirable than a dynamic evaluation (films of graphics simulation).

DRIVER-CAR RESPONSE

Three types of studies were found which examined driver-car response: simulation studies, field studies with an instrumented car and field studies where observers collected data on unsuspecting drivers.

Allen et al. (1977) used a car simulator discussed in the previous section to assess driver response to various delineation treatments under normal and degraded visibility conditions. The driver-car response measures included standard deviation of lane position, average speed, and car-driver transfer functions. This last measure is a very sophisticated measure of steering control, requiring an understanding of control theory and Fourier analysis. Because of its complexity it will not be discussed further. One result of this study was that tracking performance became erratic in fog where the gap length between dashes was such that one dash disappeared below the hoodline before a succeeding dash was visible through the fog. This is certainly an effect which would be difficult to pick up using a static technique such as slides.

Teply and Nelson (1979) used a simpler (and less expensive) driving simulator to test car-driver response. Subjects viewed video tapes of lane markings in merging and diverging areas taken from a moving car. They had a steering wheel, accelerator, and brake to simulate the car while they

viewed the video tape. They indicated to the experimenter when they would merge or diverge as the case might be.

Allen and his colleagues validated the simulator work using an instrumented car. Subjects drove a highway route at night on several occasions, with degraded delineation both before and after repainting. Four different levels of striping contrast were found along the highway. While both lane deviation and speed were significantly affected by delineation treatment in the simulator, only lane deviation was significantly effected by tests with the instrumented car. Although an instrumented car has higher face validity than a simulator, it can be less sensitive to changes in performance. The reason for this is that the driving task on a road is less defined than it is in a simulator -- particularly one such as this with a very simple graphics display. The less defined nature of the real driving task leads to a greater variability in performance and thus reduced sensitivity to differences between test conditions. Allen and his colleagues raised the interesting question -- if you improve delineation under reduced visibility conditions, will the driver then, inappropriately, increase his speed? This is another consideration in evaluating lane markings that calls for dynamic rather that static assessment.

Teply and Nelson (1979) also validated their simulator work using two types of field study. In the first, trajectories of cars on experimental merging and diverging lanes were videotaped and digitized using sophisticated computer software and hardware. The uniformity of car movement in speed and lateral positioning was measured. In the second field study observers counted critical incidents (e.g. erratic braking or lane changing) with different lane marking treatments. The results of these two field studies were in agreement with the laboratory simulation study.

Nedas, Balcar and Macy (1981) also used an instrumented car with a lane tracker to assess wide edgelines as an alcohol countermeasure. They conducted a closed-course field study at night using 4 edgeline widths and 3 levels of alcohol. Mean and standard deviation of lane position were measured. Wider edgelines were found to significantly reduce lane position variability.

SECTION IV: EVALUATION OF DELINEATORS

INTRODUCTION

The last category of highway information systems to be reviewed is delineators. The term delineator is used here to refer to pavement messages, panel and barricade forms used to direct traffic around construction, reflective pavement markers and road side reflective posts. Of all the highway visual information systems that were reviewed, delineators have received the least attention in terms of evaluation of driver response.

LEGIBILITY DISTANCE

As was true for lane markings, visibility of reflective posts and pavement markers has primarily been assessed using photometric means, despite a possible lack of correlation between photometric measures and drivers' perceptual judgements (Cole and Jenkins, 1982).

Two references were found which discussed legibility distance of pavement messages. In both studies legibility was assessed in the field and in the laboratory (MacDonald and Hoffman, 1973; Hoffman, Young and MacDonald, 1974). In the field letters were painted on an unused runway. The subject sat in the driver's seat and observed the letters through a slit one metre above ground level. The side of the runway was marked in increments of visual angle subtended by the letters. The experimenter started at the most distant mark and drove towards the letter, stopping at each mark where the viewing slit was uncovered and the subject attempted to recognize the letter. In the laboratory, slides of letters were used and presented at the same angle as subjects would view pavement messages from a car. Minutes of subtended vertical angle was the measure of legibility distance. Laboratory and field measures were found to be strongly correlated with letters in the field having generally longer legibility distance. (This was also true for the field - moving picture comparison of legibility distance for static signs by Desrosiers (1965)).

COMPREHENSION

Godthelp and Riemersma (1982) reported laboratory experiments in which drivers had to estimate the geometry of the road ahead. Subjects viewed slides taken at night of various road configurations. Four delineation systems were evaluated in the first study: panel-mounted lamps at two different heights above the roadway, post-mounted reflectors and raised pavement markers. One or two lights were removed

in each picture to better simulate conditions which occur in reality because of defects, etc. The performance measure was percentage of errors in estimating road geometry. It was found that percentage of errors was as much as 70 % for reflectors at eye height. In the second study stripe designs for barricades were evaluated using the same evaluation technique. Again devices were set out with some removed to better simulate the normal road environment. The results indicated that the use of chevrons on panels would be especially effective in conditions where panels are placed somewhat haphazardly or when a limited number of panels are available. Thus real-world disorder can interact with barricade design, i.e., the most effective barricade in an orderly display may not be the best barricade where some devices are missing or placed poorly.

PROCESSING TIME AND CONSPICUITY

Pain and Knapp (1979) performed a series of laboratory experiments to examine the optimum design characteristics of channelizing devices i.e. panel and barricade forms painted with stripes or chevrons. Subjects saw slides which were presented for 0.4 or 0.8 seconds. One of the four quadrants of the slide contained a striped bar or panel; the other three quadrants provided a visually noisy and abstract background. Subjects were required to search the slide, identify the quadrant the bar or panel was in, identify the type of design (horizontal, vertical, diagonal, chevron) and the shape (bar or panel). Correct responses (3 maximum) for each slide were summed and used as an index of performance.

Triggs et al. (1979) used a reaction time measure to assess processing rate. This study, which was discussed in the section on lane markings, examined how clearly road-side reflective post delineators gave drivers information about the direction of the curve ahead. Stimulus material for the subjects were slides taken at night showing right and left curves. Some experimentation was required to find the camera exposure time which would yield photographs corresponding to what could be seen by the driver (1 second was used). Subjects viewed the slides and pressed a button as soon as they could tell which direction the curve was in. Road contour was found to have a significant effect on discrimination speed and accuracy with hill crests yielding the poorest performance overall.

The advantage of the technique reported in the second study was that delineation was assessed in context whereas in Pain and Knapp's study an abstract background for the panels was used. Their study also combined a measure of processing rate with a measure of conspicuity and by the scoring procedure gave equal weight to both aspects of response.

This would be appropriate if the panels and barricades had been photographed in context. However the particular background used may have affected conspicuity, and thus the scores may be biased.

DRIVER-CAR RESPONSE

Allen et al. (1977) examined the effects of raised pavement edge markers on lane position control using both a graphics-based simulator and an instrumented car. In the simulator, edge markers were shown by a short dash (the reflective quality could not be simulated). Markers were evaluated under normal (90 m.) and reduced (15 m.) visibility conditions. One of the interesting findings of the simulator experiments was the interaction between visibility and delineation treatment. As discussed in the previous section delineation gap length had more effect on steering performance in reduced visibility than under normal conditions. The authors point out that even if one dash is visible it can give some curvature information, however raised pavement markers do not provide any inherent curvature information unless more than one element is visible . Lane position variability was recorded for edge striping with and without pavement edge markers.

A field study using a car instrumented with a lane tracker was then carried out. Again lane control was recorded for edge striping, with and without reflective pavement edge markers, under both wet and dry conditions. Unfortunately, the driving route was such that the combination stripes and markers were on a much straighter portion or road than the stripes only, making comparisons somewhat unreliable. Another drawback of these studies was that only six subjects participated.

Davis (1983) reported five before-and-after field studies to determine delineation needs around construction sites. Reflectors were compared with steady-burn lights, vertical panels (as recommended by Pain and Knapp's 1979 laboratory study) with bulky low barricades; raised pavement markers were compared with and without paint striping. All comparisons were made during both day and night conditions. Observers took forty-five minute samples of behaviour, counting lane changes, lane encroachments (wheels crossing line) and total traffic volume. Observers also estimated vehicle speeds; the method of doing so is not reported. Not all driver performance measures were made in each study and no reason is given for the selection of the measures that were used. As mentioned earlier Pain and Knapp (1979), in a study of driver processing rate using glance legibility found tall vertical stripes to be clearly superior to other barricade designs. An interesting result of Davis' (1983) study was that the tall panel did not appear to be significantly different from the traditional barricade in

terms of driver performance measures. However it should be noted that the performance measure used -- lane encroachment -- is fairly crude compared to the type of measures Pain and Knapp used.

CHAPTER 3: RECOMMENDATIONS

GENERAL CONSIDERATIONS IN EVALUATING HIGHWAY INFORMATION SYSTEMS

During the course of reviewing the literature on evaluation of signs, lane markings and delineators, a number of issues have arisen which apply to all types of evaluation. Prior to making specific recommendations on techniques the following general issues will be discussed: adequate sampling of the driver population, assessment under degraded visual conditions, the need for using a loading task, the need for using dynamic rather than static techniques and the setting of minimum standards.

Sampling of the Driver Population

Any assessment of visibility must consider the range of visual capabilities of the driving population. Johnston et al. (1976) discuss the 'real observer.' They argue that a serious deficiency of both field observations and laboratory investigations is that subjects in the studies are not representative of the real observer population. They point out that 27% of the adult population have corrected visual acuity which is less than normal (6/6 or 20/20), that between 2% and 10% of the population have problems detecting low contrasts at night, that the threshold for detection of low contrast stimuli increases by a factor of about 3 by the age of 60, and finally, that susceptibility to disability glare increases with age. Neither contrast threshold nor susceptibility to disability glare is tested in a driver examination, nor in most cases is there a periodic review of acuity, despite the initial requirement for a minimum of 6/12 acuity for an Ontario driver's license.

Sivak et al. (1981) showed that legibility distances at night for subjects over age 61 were 65-77% of those of the younger subjects. Hulbert and Fowler (1980) found older drivers consistently to score lower in understanding and recognition of signs and markings. Woltman et al. (1984) in an experiment on sign conspicuity, found that older subjects had more difficulty counting numbers of signs presented in short film scenes than did younger subjects. Consequently any evaluation procedure should test both older and younger subjects.

In an ethnically diverse society such as Ontario, it would also be wise to sample subjects from different cultural backgrounds, as well as from different age groups, as this might have a bearing on comprehension of signs.

Another aspect of subject selection is the use of

adequate numbers of subjects. Some of the papers reviewed in Chapter 2 used as few as 6 subjects. If the results are to be applied to a diverse population, 100 or more subjects may be required, depending on the selection procedure for the subjects, the evaluation being carried out, and the power of the statistical test. For example, if the detection of small differences between signs is desired then more subjects will have to be used than if only large differences are of interest. Appendix B discusses statistical considerations.

Assessment Under Degraded Conditions

Where the possibility exists that an interaction will occur between types of systems being evaluated and degraded visibility, the assessment should be carried out under degraded conditions. For example, Dewar and Ells (1977) assessed verbal and symbolic signs under normal contrast conditions and under lowered contrast conditions with glare present. They found that processing time for verbal signs was much more affected by the visual degradation than it was for symbolic signs.

Use of Loading Tasks

Forbes (1964) recommended that any evaluation of a sign should be carried out with a loading task because the driving situation is such that one's attention is divided between car control and monitoring the environment. (In the field the loading task can be driving.) Bhise and Rockwell (1973), in studies of driver eye-position, showed conclusively that the driver fixates several times on a sign, and that the amount of time he watches the sign is dependent on the demands of the driving task. Thus, for example, a changeable message sign with a three part message may have an acceptable legibility distance, but only if the driver does not have to perform the driving task concurrently.

Intuitively it would seem that a pursuit tracking task using rate dynamics would be the best approximation to car steering and would be the ideal loading task. Unfortunately, very little work on sign or delineation evaluation has used loading tasks.

Dynamic vs. Static Assessment

The driving environment is a dynamic one. Since the driver normally is moving while he reads a sign or follows a lane marking, it would seem preferable to carry out evaluations in a dynamic setting, using field studies, films or graphic simulation. For example, as was pointed out in the section on legibility distance, if the subject is moving toward the sign (as a car would), processing time is a factor in legibility distance. Signs that require a longer time to comprehend will have shorter legibility distances. Processing time is a component of legibility distance in a

dynamic situation only. It is particularly a factor for longer messages which can take several seconds to read. The importance of the dynamic nature of driving is underscored by Burg (1967), who showed there was a much higher correlation between dynamic visual acuity and accident rate than between static acuity and accidents.

Setting of Minimum Standards

Minimum standards must be set for any highway information system characteristic being evaluated. For example, it is not satisfactory to simply accept the most comprehensible version of a sign if average percent correct is 50%. The acceptable level will be a function of the importance of the sign. A 'do not enter one-way' sign should be more comprehensible than a 'tourist information' sign. Also, as Cairney (1982) points out, there is a need to specify minimum differences in scores which can be considered to have practical significance. With group sizes of 40 Cairney suggests that differences in percent correct comprehension of 5% or less do not have practical significance. (If the group size is smaller, a larger percent would be considered significant and vice-versa). Similarly for the ranking data, he suggests a difference of more than 0.6 in mean ranks, and for the rating data a difference of at least 0.5 for practical significance. Where difference in comprehension between two signs is small, one might select the sign with the better legibility distance. In another situation, if a new symbol is only 5% better comprehended than an old symbol, it would not be cost-effective to replace all the old signs. The setting of minimum standards is of necessity somewhat arbitrary, and is determined by practical constraints.

EVALUATION

Recommendations for evaluation of legibility, comprehension, processing rate, conspicuity and car-driver response will be made below. In the interests of simplifying the discussion, all four highway information systems will be discussed together under each driver response characteristic. After each recommended method a reference will be given to a paper discussed in the literature review in which the technique was described. Table 2 summarizes the main recommendations.

In reviewing the literature it became clear that many of the assessment methods discussed could have been administered using a combination minicomputer, software graphics and video recorder system. The capabilities of such a test facility will be discussed at the end of this report.

Table 2: Recommended Evaluation Techniques

	Static Signs	Dynamic Signs	Lane Markings	Delineators
Legibility Distance	← Reduced sign size → driver moves toward sign or vice versa		← no recommendations can be made →	
Comprehension	← preference ranking for more than 4 signs → recognition test (symbols) multiple choice recognition test (verbal messages)		← short filmed sequences with multiple choice questionnaire →	
Processing Time	← reaction time →			
Conspicuity	features seen in 1/2 second exposure of slide	not necessary	not necessary	no recommendations can be made
Driver-Car Response	not recommended (insensitive)	field study* exit volumes for accident diversion license plate recording for special events and construction diversion	← not recommended (impractical, insensitive) →	

*not possible to implement on computerized test facility

LEGIBILITY DISTANCE TESTING

Static and Dynamic Signs

Because of the dynamic nature of the driving task, a dynamic test of legibility distance is strongly advised. Dynamic testing will eliminate the need for additional testing of driver processing rate. As discussed in the literature review, processing rate comes into play if legibility distance is evaluated with the subject and sign moving relative to one another. Processing rate is a larger factor in legibility distance for verbal messages than for symbolic messages. Processing rate would be of particular importance in evaluating changeable message signs with messages in several parts.

In absence of simulation equipment, a field study using reduced size signs is recommended (Dewar and Ells, 1974b). The reduction in sign size allows the driving to be carried out at lower speeds and thus decreases the length of road required for testing. MTC's own driver licensing facility might be considered as a test site. Because the driving task forces the driver to read a sign in several glances, only drivers, not passengers should be used as subjects. Each driver should approach the sign four times, twice from each of two directions, so that differences in background lighting will not bias results. Testing should be carried out under cloudy and under sunny conditions and at night; changes in background have been shown to affect legibility distance by as much as 25%, also night conditions have been shown to reduce greatly the legibility distance for older drivers.

If a computerized test facility were available, the symbol or verbal message could be presented using graphics software. The symbol or message tested should increase in size until the subject can read it. The rate of increase in size should simulate the speed at which the subject would normally approach the sign. A standard sign (or signs) whose legibility distance in the field is known should be tested along with the signs of interest so that equivalent field legibility distance can be calculated. Dewar and Ells (1974b) and Desrosiers (1965) showed that legibility distances obtained using a dynamic technique in the laboratory (film or subjects being moved in wheelchairs towards a slide) were similar to those obtained in the field. However, due to the greater luminance and greater contrast of the field situation, distances measured in the laboratory were shorter. Therefore, some standard signs are required against which signs of interest can be measured. Testing under various background lighting conditions could be simulated by changing the screen luminance.

Lane Markings and Delineators

Legibility of lane markings and delineators is of particular concern in adverse environmental conditions. Lane markings are particularly problematic in rain. With delineators dirt is a problem; smaller delineators or delineators to which dirt adheres easily may suffer greater loss in legibility than larger delineators. Thus legibility distance of lane markings and delineators should both be tested under field conditions. Dirt and rain are difficult to simulate well.

Unfortunately no studies were found in the literature in which drivers assessed legibility distance. In one study, the experimenter estimated legibility distance by counting the number of visible marks and multiplying marks by the delineation gap plus mark length. Otherwise, visibility was estimated by photometric means, despite convincing evidence that perceptual judgements are not always linearly related to physical measures. Further research on techniques is needed before an evaluation method for legibility distance can be recommended.

Special Considerations

If interaction is expected between degraded visibility and the signs or markings being compared, then testing should also include degraded visibility. For example, symbols are not as affected by a decrease in contrast as verbal messages. Also the gap between lane markings affects legibility distance during fog. In both these situations testing under degraded conditions would be required. Simulation of reduced contrast due to fog is relatively simple with a graphics presentation -- certainly simpler than a technique used by Hodge and Rutley (1978). These researchers simulated fog by boiling vegetable oil and feeding the steam into the laboratory ventilating system! It is also difficult to study fog in field conditions. Research teams have to stand by for indefinite periods of time and when they do arrive at a site with fog, the fog can disappear very quickly.

COMPREHENSION TESTING

Subject selection is of particular importance in comprehension testing. Subjects should be representative of the Ontario population in age, sex, education and cultural background and driving experience.

Static and Dynamic Signs

Where four or more alternate symbols or verbal formats are to be evaluated for the same sign message, they should be preference ranked according to which symbol or message communicates the intended meaning best. A well selected

sample (see introduction) of subjects should be used.

From the preference ranking procedure 3 symbols or messages should be selected for further evaluation using the recognition method for symbols and a multiple choice recognition test for verbal messages (Dewar and Ells, 1974). The recognition method, which is time consuming to score, is recommended for symbols because of the valuable information wrong interpretations may give. For example, with especially important signs (e.g. one way street), a finding that 10% of subjects interpreted the symbol to have the opposite meaning would suggest a redesign of the sign or an intense education campaign. Such wrong interpretations are not so readily revealed using the multiple choice technique. Verbal signs have fewer possible meanings in most cases than do symbol signs, thus they are more appropriately tested using a multiple choice recognition test. Selection of good alternative answers for the multiple choice test is critical. Fifteen to twenty subjects should be asked the meaning of the symbol(s) being tested (the recognition method). Their wrong answers and other meanings they suggest should then be used to generate the multiple choice test questions.

Lane Markings and Delineators

Comprehension of lane markings and delineators should be evaluated using short filmed sequences interspersed with multiple choice questionnaires asking the subject which lanes he or she is permitted to travel in, and in the case of lane markings, whether passing or turning in the adjacent lane is permitted (Gordon, 1976). Questions should be phrased so as to have simple answers: yes, no, maybe, left, right. Hulbert and Fowler (1980) used special ink and paper so that when the subject marked the answer, he or she saw immediately if the answer is correct. Thus subjects were educated at the same time they were tested. This could be considered only if learning effects were carefully controlled by presenting the questions in different orders, which may not always be practical. Films can be presented to large groups so that the data can be collected efficiently. If a computerized test facility were available, film sequences could be presented on videotape and the questionnaire presented on the terminal screen and the keyboard could be used for subject responses. Scoring would then be automated.

Special Considerations

New Symbols Similar to Current Symbols: If a new symbol being introduced is similar to other symbols already in use, the degree of confusability with the current symbol set should be tested using the matching technique (Dewar and Ells, 1974a) where 3 possible symbols to convey a particular message are presented to the subject. The subject selects the symbol that best conveys the message. Each set should include any symbols in current use which may be confused with

a new symbol. This is a test which is easily scored but takes some time to set up. If it were presented on a computer, scoring would be automated.

Changeable Message Sign Abbreviations: Abbreviations for changeable message signs should be tested first by having subjects generate their own abbreviations. To establish the population stereotype the method subjects used should be classified. Examples are: drop all vowels, use first three letters and last etc. A large group of subjects should be tested to see which way of generating abbreviations is most common. The recognition method may then be used to ensure the abbreviations chosen are well comprehended. In this case both scoring and test set-up would be very simple.

Learnability of Symbols: If evaluation funds permit a recall test should be administered to the subjects three weeks after they have participated in recognition testing of the symbol to see how many recall the correct meaning. If the symbol is quickly forgotten a redesign or else a more lengthy education campaign than might normally be conducted would be indicated. The subjects for this test should be drawn from the original group tested.

PROCESSING TIME TESTING

Processing time has mainly been measured using two techniques: glance legibility and reaction time. With the first technique one attempts to find a stimulus presentation time that will result in 30-60% of responses being correct. This can be time consuming; also there are no particular advantages of this technique over reaction time measures. Thus the reaction time measure is recommended in assessment of signs and lane markings (Dewar and Ells, 1974b). Pavement markings and delineators need assessment in situations where discontinuities occur. Examples for lane markings are: intersections, lane drop-offs, lane narrowing etc., and for delineators: barricades around construction sites, and curves where direction of the roadway changes. It is not necessary to assess continuous lane markings or delineators that are continuous with such a technique because the driver is not limited in the time available to process the information. Reaction time to signs, markings or delineators can be measured in the laboratory using slides or on a computer screen using graphics presentation. Processing time should be evaluated even when evaluation of legibility distance is dynamic in certain situations. For example where a changeable message sign is very long or is presented in several parts, subjects may indicate they see the sign as soon as they can read the first couple of words. For signs such as this processing time should be measured to ensure that the whole sign can be read at a sufficient distance away.

CONSPICUITY TESTING

Compared to the other driver response characteristics examined, the fewest studies were found on conspicuity. As indicated in the literature review, there were studies of conspicuity of static signs, delineators in the form of barricades, and, in a very simple way, of changeable message signs. The poverty of literature and the differing definitions of conspicuity make it difficult to recommend techniques for testing.

Conspicuity is of greatest concern for signs, markings or delineators which have serious consequences if they are missed. It is of particular concern in situations where many other stimuli compete with the signs or markings. It is only in these situations that conspicuity need be tested.

Static Signs

It is recommended that static signs be tested using the technique outlined by Cole and Jenkins (1982). Slides should be taken of signs to be tested in a variety of backgrounds, similar to the backgrounds in which they will be placed and in both daytime and nighttime conditions. Camera distance from the sign must be standardized. Subjects should view the signs for half a second after they have fixated on a central point immediately prior to seeing each slide. They should then report any road fixtures seen that attract their attention. If a large percentage of the subjects miss the sign, then many factors may be considered which would alleviate the problem, not all of which would involve redesign of the sign -- placement can be of equal or greater importance.

Dynamic Signs

Conspicuity is not thought to be a serious problem for dynamic signs which are generally in highway environments which are not visually complex. Also they do not compete with other similar signs.

Lane Markings and Delineators

In general, the road surface is uncluttered and conspicuity of lane markings and delineators is not a major issue. However, around construction sites, with lane markings being incompletely painted out and particularly in rain at night, conspicuity of markings and delineator barricades or panels can be a serious problem. The old lane markings can be clearer to the driver than the new ones, particularly if the driver is accustomed to the road as it was before construction. Another situation where conspicuity is an issue is that of viewing white lane markings on snow covered roads. It is hard to imagine an adequate simulation

of either of these situations. A field test using observers to record lane encroachments in daytime, nighttime, snowy and rainy conditions would seem to be a good test of conspicuity. Unfortunately no studies were found on which this recommendation could be based.

DRIVER-CAR RESPONSE TESTING

Static Signs

The use of driver-car response measures such as speed changes and lane positioning to test static signs was found to be unsuccessful. Laboratory and field studies of sign characteristics such as legibility distance and comprehension were more sensitive in distinguishing the best sign to select. In any case, many signs alert a driver but do not evoke an overt and therefore measurable reaction e.g. warning signs. Driver-car response testing is not recommended in static sign evaluation.

Dynamic Signs

In contrast to the situation with static signs, driver-car response measures are vitally important in assessing dynamic signs. The main reason for the use of changeable message and symbol signs is to affect traffic flow, i.e. to divert traffic to other lanes or other roads, or to slow traffic down. Consequently any evaluation of dynamic signs, particularly changeable message signs, would not be complete wihout an assessment of driver-car response. The recommended technique is dependent on allowable expenditure for the evaluation as well as on the type of message being evaluated.

For messages about predictable events such as construction or special events, the method used by the Texas Transportation Institute and reported by Weaver et al. (1977) is recommended. The messages should be displayed for short (say 15 minutes) intervals, interspersed with blank intervals of equal length. Numbers of vehicles exiting the highway after the message sign as well as numbers of vehicles remaining on the highway should be recorded. License plate numbers should be recorded at all points to ensure complete origin-destination information (and also to avoid including drivers who enter the highway beyond the sign location and exit immediately). A car which floats with the traffic stream should be used to classify drivers as to whether they saw a message or not and which message they saw. It is also recommended that license numbers be used to obtain driver addresses and a questionnaire be sent out as quickly as possible (less than a week) to drivers seeing the signs. Questionnaires would help determine why drivers responded to the signs the way they did. To ensure a high return rate, questionnaires should be brief and clear. The lower the rate

of return, the more possibility exists that the questionnaire results will be biased.

The type of evaluation recommended above is very thorough but also expensive. Numerous people are required to conduct the evaluation, the development of a good questionnaire is time consuming as is the obtaining of driver addresses from license numbers. If funds allow, the type of evaluation reported by Hanscom (1981) where drivers were recruited to drive past test sites and then answer questionnaires is recommended as a technique of validating questionnaire answers about response to signs. Hanscom reported that the behaviour of the recruited drivers was the same as the unobtrusively observed drivers. However, by recruiting drivers you remove selection bias to a great extent. There is not the possibility that the questionnaire results would be biased as they can be if the return of the questionnaire is optional. Also the driving sample can be selected to be representative of the Ontario driving population in age, sex, educational and cultural background.

In evaluating messages about traffic tie-ups due to unpredictable events, such as accidents, the technique described by Dudek et al. (1982) is recommended. Traffic counters were placed on the freeway, on interchange ramps and on primary off-ramps. The counters were modified to produce 5-minute traffic counts and feed this information directly to a central computer. Alternately traffic counts could be recorded on paper tape along with the time of the count. Normal traffic flow patterns were recorded as well as traffic flow patterns after an accident. 'Natural' diversion occurs when drivers see the traffic slowing in advance of seeing what is causing the tie-up. Traffic flow was assessed both with the changeable message sign operating during an incident and not operating to estimate the amount of 'natural' diversion as opposed to diversion because of sign operation. The problems encountered by Dudek et al. (1982) underscore the necessity of having a 'closed system' so that all traffic inputs and outputs from the incident area are accounted for. Though Dudek et al. (1982) did not report the use of a questionnaire as part of their evaluation, it is recommended that a questionnaire be used to augment the information obtained from a study of traffic flow. Continuous-loop video tape recording of license numbers might be considered to identify drivers affected by a particular traffic incident.

As will be discussed later a computerized test facility would allow testing of driver reaction to changeable message signs by presenting a subject with a video-game type simulation where he or she sees changeable message signs and can elect to stay on the highway or divert with the appropriate consequences in time delay for each action. Such a simulation would produce useful information but a field test would still be required.

Lane Markings and Delineators

Driver-car responses are important for evaluating some aspects of lane markings and delineators. For example width of markings, gap distance between markings in fog and placement of delineators relative to the road edge are all likely to affect lane positioning behaviour. However, the lane positioning behaviour is likely to be affected in very subtle ways requiring either the use of an instrumented car with a lane tracker to evaluate, or the counting of numbers of lane exceedances for hundreds of cars. The first technique is very expensive. The second technique is feasible only where large traffic volumes exist. Allen et al. (1977) showed driver workload, as estimated using a subjective rating scale, increased before changes in lane position became evident. For this reason and for reasons of practically it is recommended that other characteristics of lane markings and delineators be assessed, such as comprehension and legibility distance, rather than driver-car responses.

TEST FACILITY

Almost all the evaluation procedures dicussed above could be carried out using a combination microcomputer, software graphics and video recorder system.

There are numerous other advantages, besides the flexibility discussed above, of a computer graphics/video tape recorder system. First, it can be designed to be portable so that testing can take place at different locations. At one location a group may consist mainly of older drivers, at another (e.g. a high school) the group may consist of younger drivers. Second, testing would be conducted efficiently; if tests were presented one after the other the same subjects could be used to assess legibility distance, comprehension etc. Having subjects do several different evaluations also saves recruiting subjects over and over again. Thirdly, scoring of tests is automated. Even recognition test scoring would be simplified if answers were keyed in and the examiner could score them at the computer. Data collected on floppies could easily be transferred to a larger computer for statistical analysis. Fourth, dynamic testing is possible -- signs and lane markings can be seen in motion, just as in the real world. The possibility of adding a loading task, preferably a tracking task, adds to the realism of the test situation. Fifth and finally, fabrication costs of signs and costs of painting roads with proposed markings would be reduced by having graphics simulation available for the first stages of testing.

There are three main areas of concern in the design of the test facility described above which will affect the

feasibility of using it to test driver response. These are the quality of the visual image presented both by the graphics software and by the video recorder, the speed of the processor, which affects the length of time it takes to create each new image, and the refresh rate (number of times per second the picture is redrawn) of the monitor.

With a raster scan presentation the image is made up of a matrix of dots, and the resolution of the image is limited by the size of the matrix or raster. It would be completely appropriate to test dynamic signs using even a low resolution image because the signs themselves are made from a matrix of lights, analagous to the raster scan of a monitor. However static signs, particularly static signs showing symbols, require high resolution to present accurately. The higher the resolution required the more expensive the system. High resolution of the static signs will certainly be required for legibility, processing time and conspicuity testing. Zooming a graphics image towards a subject to test legibility will present distance problems if the processor is not fast enough to redraw successive images at the required speed. Terminals are available which attach to mainframe computers and use specialized hardware to allow zooming and panning of the image. Currently these are in the $30,000 range. Zooming may have to be achieved by writing special software or by filming successively larger graphic images using a camera which can record a single frame at a time.

The refresh rate of the monitor is of concern because it places a limit of the presentation time for the image. Most monitors are refreshed 60 times per second, thus the smallest presentation time would be 16 msec. and would increase in increments of 16 msec. This is not thought to be a serious limitation however. Testing of information processing rate has been carried out successfully at McMaster University and at Sunnybrooke Hospital using a computer presentation.

Woltman et al. (1984) successfully tested conspicuity of static signs in a simple environment using a video recorder image. However neither legibility nor processing time for static signs has been tested using a computer graphics or a video recorder image. Just as all other methods recommended for evaluation of highway systems have been tested, it would be necessary to ensure that the visual image produced by the graphics software and the video recorder or the monitor, as well as the refresh rate, were adequate for testing legibility and processing time.

If an acceptable method of zooming the image for legibility testing can be worked out, without requiring expensive zoom and pan features, it should be possible to purchase a computer, high resolution monitor, video recorder and camera, A/D converter (to connect the computer and recorder) for under $15,000.

REFERENCES

Allen, R.W., O'Hanlon, J.F. and McRuer, D.T. Driver's Visibility Requirements for Roadway Delineation Vol. I Effects of Contrast and Configuration on Driver Performance and Behaviour. Dept. of Transportation, FHWA/RD-77/165, Nov. 1977. A45 *

Bhise, V.D. and Rockwell, T.H., Toward the Development of a Methodology for Evaluating Highway Signs Based on Driver Information Acquisition, Highway Research Board Record, 1973, 440, 38-56. A1

Burg, A. The relationship between vision test scores and driving record: General findings. Report Number64-24, Los Angeles, University of California, Department of Engineering, June 1967.

Burg, A. and Hulbert, S.F., Predicting the effectiveness of highway signs, Highway Research Board Bulletin, 1960, 324, 1-11. A2

Cairney, P.T., Validity of Preference Ranking as a Screening Procedure for Symbolic Signs, ARRB Internal Report, AIR 1109-3, September, 1982. A3

Cole, B.L. and Jenkins, S.E. Conspicuity of traffic devices. Australian Road Research, 1982, 12(4), 223-238. A4

Davis, T.D. Construction Zone Safety and Delineation Study, N.J. Dept of Transportation, FHWA/NJ-83/005, February, 1983. A49

Desrosiers, R.D., Moving Picture Technique for Highway Signing Studies -- An Investigation of Its Applicability, Public Roads, 1965, 33(7), 143-147. A5

Dewar, R., Static Traffic Signs: A State-of-the-Art Review, presented at the International Symposium on Traffic Control Systems. Berkeley, California, August 6-9, 1979. A6

* annotated bibliography page number

Dewar, R.E. and Ells, J.G., Methods for the Evaluation of Traffic Signs (Final Report), Ministry of Transport, Ottawa, Contract 97096, June 1974, a. A7

Dewar, R. E., and Ells, J.G., Comparison of Three Methods for Evaluating Traffic Signs, TRB/TRR, 1974, b, 503, 38-47. A8

Dewar, R.E. and Ells, J.G., The Semantic Differential as an Index of Traffic Sign Perception and Comprehension, Human Factors, 1977, 19(2), 183-189. A9

Dewar, R.E., Ells, T.G. and Mundy, G., Reaction Time as an Index of Traffic Sign Perception, Human Factors, 1976, 18(4), 381-392. A10

Dewar, R.E. and Swanson, H.A. Recognition of traffic control signs. Highway Research Record, 1972, 414, 16-23.

Dietrich, C.W., Markowitz, J. and Jonew, C. Investigation of New Traffic Signs, Markings and Signals. Vol. 1: Laboratory Experiments and Road Tests, Vol. 2: Driver Questionnaire, Bolt, Beranek, and Newman Inc. December, 1972.

Dudek, C.L., Messer,C.J. and Jones, H.B. Study of Design Consideration for Real-Time Freeway Information Systems. Highway Research Record, #363, p. 1-14, 1971. A28

Dudek, C.L., Stockton, W.R. and Hatcher D.R. Real-time Freeway-to-Freeway Diversion -- The San Antonio Experience, Texas Transportation Institute, January 1982 (draft). A29

Dudek, C.L., Weaver, G.D., Hatcher, D.R. and Richards, S.H. Field Evaluation of Messages for Real-Time Diversion of Freeway Traffic for Special Events, Transportation Research Board Record #682, 37-45. A30

Easterby, R.S. and Zwaga, H.J.G. Evaluation of public information symbols -- ISO tests: 1975 series. AP report 60. Applied Psychology Dept., University of Aston, Birmingham, March 1976.

Ells, J. G. and Dewar, R.E., Rapid Comprehension of Verbal and Symbolic Traffic Sign Messages. Human Factors, 1979, 21(2), 161-168. A11

Forbes, T.W., Fry, J.P., Joyce, R.P. and Pain, R.F. Letters and sign contrast, brightness, and size effects on visibility. Highway Research Record, 1968, 216, p. 48. A12

Forbes, T.W., Fry, J.P., Joyce, R.P., and Pain, R.F. Colour and brightness factors in simulated and full-scale traffic sign visibility. Highway Research Record, 1968, 216, p. 55.

Forbes, T.W. and Holmes, R.S. Legibility distances of highway destination signs in relation to letter height. Highway Research Board Proceedings, 19, 1939, 321-335.

Forbes, T.W., Moskowitz, K. and Morgan, G. A Comparison of Lower Case and Capital Letters for Highway Signs. Proceedings of Highway Research Board, 1950, 30, 355-373.

Forbes, T.W., Snyder, T.E., and Pain R.F. A Study of Traffic Sign Requirements. II Annotated Bibliography, Dept. of Psychology, Michigan State University, August, 1964.

Godthelp, J. and Riemersma, J.B.J. Vehicle guidance in road-work zones. Ergonomics, 1982, 25(10), 909-916. A50

Gordon, D.A. Studies of the Road Marking Code. Federal Highway Administration, FHWA-RD-76-59, April, 1976. A46

Gordon, D.A. Legibility of Highway Guide Signs, TRB/TRR, 1982, 855, 1-6. A13

Green, P. and Pew, R.W. Evaluating Pictographic Symbols: An Automotive Approach, Human Factors, 1978, 20(1), 103-114. A14

Hanscom, F.R. Effectiveness of Changeable Message Signing at Freeway Construction Site Lane Closures, Transportation Research Record, #844, 35-41, 1981. A31

Heathington, K.W., Worrall, R.D. and Hoff, G.C. An Analysis of Driver Preference for Alternative Visual Information Displays, Highway Research Board Record #303, 1-11, 1970. A32

Hodge, A.R. and Rutley, K.S. A Comparison of Changeable Message Signals for Motorways, TRRL Supplementary Report #380, 1978.

Hoffman, E.R., Young, P.A. and MacDonald, W.A. Recognition of road pavement messages -- a simulation study. Australian Road Research Journal, Vol. 5, No. 6, December, 1974.

Huchingson, D. and Dudek, C.L. Delay, Time Saved and Travel Time Information for Freeway Traffic Management, Transportation Research Record #722, p. 36-39. A34

Hulbert, S. and Beers, J. Research Development of the Changeable-Message Concept for Freeway Traffic Control, Highway Research Board Special Report, #129, 1971, 24-25. A35

Hulbert, S. and Beers, J. Wrong way driving: off-ramp studies, Highway Research Records, 1966, 122, 35-49. A15

Hulbert, S., Beers, J. and Fowler, P. Motorists' Understanding of Traffic Control Devices. AAA Foundation for Traffic Safety, March 1979. A16

Hulbert, S. and Fowler, P. Motorists' Understanding of Traffic Control Devices: Test II, AAA Foundation for Traffic Safety, December, 1980.

Jacobs, R.J., Johnston, A.W. and Cole, B.L. Visibility of alphabetic and symbolic traffic signs, Australian Road Research, 1975, 5(7), 68-86. A17

Johansson, G. and Backlund, F. Drivers and road signs. Ergonomics, 1970, 13, 749-759. A18

Johnston, A.W., Cole, B.L., Jacobs, R.J. and Gibson, A.J. Visibility of Traffic Control Devices: Catering for the Real Observer. Ergonomics, 1976, 19(5), 591-609 A19

Knapp, B.G. and Pain R.F. Human Factors Considerations in Arrow-Board Design and Operation, TRB/TRR #703, p. 1-8, 1979. A36

MacDonald, W.A. and Hoffman, E.R. The recognition of road pavement messages. Journal of Applied Psychology, Vol. 57, No, 3, 1973.

Mackett-Stout, J. and Dewar, R.E. Evaluation of Symbolic Public Information Signs, Human Factors, 1981, 23(2), 139-151. A20

Markowitz, J., Dietrich, C.W., Lees, W.J. and Farman, M. An Investigation of the Design and Performance of Traffic Control Devices. Bolt Beranek and Newmnan Inc., Report No. 1726, Contract No. CPR-11-5955, 1968.

McNees, R.W. and Messer C.J. Reading Time and Accuracy of Response to Simulated Urban Freeway Guide Signs, Texas Transportation Institute, January 1982. A21

Mounce, J.M., Messer, C.J., Huchingson, R.D. and Dudek, C.L. Bulb-loss Effects on Message Readability of Motorist Information Matrix Signs. TRB/TRR #643, p. 40-45, 1977. A37

Mourant, R.R., Rockwell, T.H., and Rackoff, N.J. Drivers' eye movements and visual workload. Highway Research Record, 1969, 292, 1-10.

Nedas, N.D., Balcar, G.P., and Macy, P.R. Road Markings as an Alcohol Countermeasure in Traffic Safety: A Field Test of Standard and Wide Edgelines. Presented at the Annual Meeting of the Transportation Research Board, January 14, 1981. A47

Noble, M. and Sanders, A.F. Searching for Traffic Signals while Engaged in Compensatory Tracking. Human Factors, 1980, 22(1), 89-102. A22

Pain, R.F. and Knapp, B.G. Experimental Evaluation of Markings for Barricades and Channelizing Devices. TRB/TRR, 1979, 703, 8-16. A51

Osgood, C., Suci, G. and Tannenbaum, P. The measurement of meaning. Urbana, Ill.: University of Illinois Press, 1957.

Richards, S.H., Stockton, W.R. and Dudek, C.L. Analysis of Driver Responses to Point Diversion for Special Events, TRB/TRR #682, p. 46-52, 1978. A38

Roberts, K.M., Lareau, E.W. and Welch, D. Perceptual Factors and Meanings of Symbolic Information Elements: Volume II, Technical Report. Federal Highway Administration, FHWA-RD-77-65, June, 1977.

Rutley, K.S. and Webb, P.J. Recognition distances and understanding of legends on an experimental motorway signal. TRRL Supplementary Report 500, 1979. A39

Seymour, W.H., Deen, R.C. and Havens, J.H. Traffic Control for Maintenance on High Speed Highways. Final Report. Research Report 327, Kentucky Dept. of Highways, Division of Research, 1972. A23

Sivak, M., Olson, P.L. and Pastalon, L.A. Effect of Driver's Age on Nighttime Legibility of Highway Signs, Human Factors, 1981, 23(1), 59-64. A24

Snyder, H.L. and Taylor, G.B. The Sensitivity of Response Measures of Alphanumeric Legibility to Variations in Dot Matrix Display Parameters. Human Factors, 21(4), 457-471, 1979. A40

Stephens, B.W. Some Principles for Communicating with Drivers Through the Use of Variable-Message Displays, Highway Research Board Special Report #129, 1979, p. 2-6. A41

Stockton, W.R., Dudek, C.L., Fambro, D.B. and Messer, C.J. Evaluation of a Changeable Message Sign System on the Inbound Gulf Freeway, Texas Transportation Institute, Research Report 200-IF, 1975. A25

Summala, H. and Naatanen, R. Perception of highway traffic signs and motivation. Journal of Safety Research1974, 6(4), 150-154.

Teply, S. and Nelson, T. Evaluation of Pavement Markings for Merging and Diverging Areas. Report to Transport Canada, 1979.

Testin, F.J. and Dewar, R.E. Divided Attention in a Reaction Time Index of Traffic Sign Perception, Ergonomics, 1981, 24(2), 111-124. A26

Triggs, T.J., Harris, W.G., and Fildes, B.N. Delineation Cues on Rural Roads at Night: A Laboratory-Based Study of Curve Direction Estimation. Australian Road Research Board, AIR-226-2, July 1979. A48

Turner, J.M., Dudek, C.L. and Carvell, J.D. Real-Time Diversion of Freeway Traffic During Maintenance Operaions. TRB/TRR #683, 1978, p. 8-10. A43

Weaver, G.D., Dudek, C.L., Hatcher, D.R. and Stockton, W.R. Approach to Real-Time Diversion of Freeway Traffic Events, TRB/TRR, 1977, 664, 57-61. A44

Woltman, H.L., Stanton, R.H., and Stearns R.A. Information Sign Color Evaluation Employing a Video Presentation, Traffic Control Materials Division/3M Audio Visual Communication, St. Paul, Minn. 55101, January 1984. A27

Zoethout, W.D. Physiological Optics, 4th ed. Chicago: Professional Press, 1947.

APPENDIX A : Annotated Bibliography

STATIC SIGNS FIELD PROCESSING TIME

Bhise, V.D. and Rockwell, T.H., Toward the Development of a Methodology for Evaluating Highway Signs Based on Driver Information Acquisition, Highway Research Board Record, 1973, 440, 38-56.

Study was conducted to develop a methodology for evaluating traffic signs by use of an eye-marker camera as a means of data collection.

Method: data was collected by using an instrumented vehicle that was equipped to measure eye movement (corneal reflection) and driving performance. Task: subject followed a route involving freeway driving having been given only entry and exit points as instructions. Measures: eye movements in relation to traffic signs and driver performance.

Results:
- in general drivers do not fixate on a sign but make several glances to it
- the time-sharing process of the drivers with the signs and other objects on the road is found to be dependent on factors such as time distance to first fixation on the sign, traffic density, type of informational need of driver, length of message displayed on the sign, relevancy of information to driver, and driver familiarity

STATIC SIGNS LAB COMPREHENSION

Burg, A. and Hulbert, S.F., Predicting the effectiveness of highway signs, Highway Research Board Bulletin, 1960, 324, 1-11.

Total of 880 subjects participated in study designed to predict the effectiveness of 5 highway signs (each warning of lane ending). Method: film-strip technique used (approx. 2 1/2 min. for each sign, 5th sign (graphic) was photographed and a lantern slide made). Subjects were tested in groups. Films were shown in random order. Task: Subjects were to fill out a questionnaire after each film. Measures: first impressions, clarity ratings, preference ratings. Results: sign "D" ranked higher on all three dimensions.

STATIC SIGNS LAB COMPREHENSION

Cairney, P.T., Validity of Preference Ranking as a Screening Procedure for Symbolic Signs, ARRB Internal Report, AIR 1109-3, September, 1982.

This study aimed to assess the ability of the ranking procedure compared with the ability of a rating procedure to predict recognition and recall of unfamiliar symbolic traffic signs. Subjects: N = 43. Method: colour prints (90 mm X 130 mm) were made of slides used in Johnston (1980) study. All versions of a particular referent were mounted in a plastic sheet, together with a panel containing a brief explanation on the sign and its possible applications. Task: (1) 25 subjects were asked to rank the versions for each referent according to how they believed them to convey the intended message (2) 18 different subjects rated each version of the referents on a 5 point scale (1 = a completely ineffective sign; 5 = a sign as effective as any ever seen before). Subjects were familiarized with material before rating. Measures: preference ranking; preference rating. Criterion measures: results of recognition and recall tests from Johnston (1980) experiment.

Results: ranking procedure performed better as a predictor of recognition scores than did the rating procedures -- selecting the wrong version on approximately 1/6 compared to 1/3 of occasions respectively.
- high proportion of successful predictions by both procedures of recall performance was apparently due to a uniformly high level of correct recall (recall test was fairly insensitive for this type of evaluation).

Cole, B.L. and Jenkins, S.E. Conspicuity of traffic devices. Australian Road Research, 1982, 12(4), 223-238.

Investigations were conducted in three parts. The first two parts estimate the probability of noticing a traffic control device in a daytime road environment. Part III attempts to determine what attributes of target and background are responsible for visual conspicuity of the target.
Part I & II 17 observers were shown 100 colour slides depicting various nominated traffic control devices in a variety of urban and semi-urban road scenes. Each slide was shown for 500 ms (Part I0 and 1500 ms (Part II). Task: Observers were asked to report the road fixtures they saw that were pertinent to them as drivers. Observers wrote their responses, indicating for each feature seen, its class, approximate location, message displayed, and any other pertinent details. Measures: number of features detected. Part III: 4 observers were shown each of the 100 slides and asked to judge the primary target and background with respect to nine subjective attributes. Measure: subjective ratings

Results: Few traffic control devices achieved a detection probability of 0.90 or more in a 500 ms observation period. Size and edge definition were important determinants of conspicuity. Brightness was less effective and white signs were noticed less frequently than coloured ones. Long observation periods were not effective in detecting additional signs.

STATIC SIGNS LAB/FIELD LEGIBILITY DISTANCE

Desrosiers, R.D., Moving Picture Technique for Highway Signing Studies -- An Investigation of Its Applicability, Public Roads, 1965, 33(7), 143-147.

A comparison of a laboratory technique and a field technique for evaluating highway signs. FIELD TEST: 28 subjects participated. Test signs were constructed at 1/3 size of real signs and vehicle speed of 20 mph was selected to simulate reading time for standard size signs at 60 mph. Test words were nonsense words embedded in list-type destination signs displaying 2, 4, or 6 legends. Task: subject drove through course at 20 mph and responded by reading out loud the selected destination and the line on which it appeared. Measure: legibility distance. LAB STUDY: 26 subjects participated. A 16 mm colour film of the 20 test signs was shown to the subjects. Task: subject was to "follow" a destination throughout the film. Responses were made by depressing the appropriate button as soon as possible for each sign (0 if test word was not present, 1 - 6 for position on sign). Measure: response time.

Results: The legibility distances obtained for different signs in the lab study were ordered in the same way as those obtained in the field, however field legibility distances were 5-6 time greater.

STATIC SIGNS REVIEW

COMPREHENSION
LEGIBILITY DISTANCE
PROCESSING TIME
CONSPICUITY

Dewar, R., Static Traffic Signs: A State-of-the-Art Review, presented at the International Symposium on Traffic Control Systems. Berkeley, California, August 6-9, 1979.

This paper reviews the major factors influencing the perception of static traffic signs and the relevant research findings regarding these factors. A number of perception and information processing variables are reviewed including: conspicuity, detection, sign legibility, nighttime perception, comprehension and the role of colour and shape in the traffic environment. Reviewed also are the research and problems associated with the use of symbols as traffic messages. A detailed examination of three types of signs -- guide signs, stop, and exclusive lane use signs -- is also included. Several recommendations made include: need for greater appreciation of driver's information needs; uniformity of sign standards; integration of sign design and placement with roadway design during the planning stages; and the need for cost-benefit analysis of traffic sign implementation.

STATIC SIGNS LAB/QUESTIONNAIRE COMPREHENSION

Dewar, R.E. and Ells, J.G., Methods for the Evaluation of Traffic Signs (Final Report), Ministry of Transport, Ottawa, Contract 97096, June 1974, a.

LAB STUDY Subjects: N = 40. Stimuli: 18 colour slides of foreign traffic symbols. Method: 1. Subjects first tested for knowledge of meaning of 18 signs presented using colour slides. 2. Each sign shown again and subject given verbal meaning. 3. Subjects give meaning to each slide shown (corrected if wrong). Slides shown until subject could identify all signs correctly on 3 consecutive trials. 20 subjects returned 3 weeks later to test memory of signs. Task: write down meaning and cue used to remember. Measures: percent errors.

Results: signs which have low intrinsic meaning are difficult to learn and more difficult to remember than signs with high intrinsic meaning.

Questionnaire Studies: A. Two questionnaire techniques were compared with a lab study. Subjects: 220 answered Quest. #1, 224 completed Quest. #2, and 60 subjects participated in lab. (visitors to an university open house.) Methods: Quest. #1 (matching) 10 traffic sign messages, with 3 black and white symbols drawn for each Task: select symbol which best conveyed specific message. Quest. #2 (multiple choice) sign drawn in black and white with 4 meanings beside it. Task: select best meaning. Lab Study 80 symbols shown, 15 seconds each. Task: write down meaning of symbol. Measure: comprehension. Results: multiple choice technique (symbol and 4 meanings) can adequately measure sign comprehension.
Exp't 14 determine comprehension of foreign traffic sign symbols using multiple choice technique. Subjects: N = 441. Stimuli: 9 foreign symbols. Method: see Quest. #2 above. Measure: comprehension.
Exp't 15 determine preference for symbols conveying specific traffic sign messages. Subjects: N = 926. Stimuli: coloured photos of traffic signs. Method: see above.
Exp't 16 4 part questionnaire. Subjects: N = 2,350. Method: coloured photos of 10 symbols, meaning typed underneath. Task: indicate on a 5-point scale (1 = very clear, 5 = very unclear) how well symbol conveyed meaning Measure: clarity of meaning.

Results (14, 15, 16): little consistency in information conveyed across these three techniques.

STATIC SIGNS FIELD/LAB COMPREHENSION
LEGIBILITY DISTANCE

Dewar, R. E., and Ells, J.G., Comparison of Three Methods for Evaluating Traffic Signs, TRB/TRR, 1974, b, 503, 38-47.

Three experiments were conducted to compare three methods of evaluating traffic sign perception.
Exp't #1: 16 subjects participated in closed-course field study. Method: 16 regulation-size traffic signs were presented as stimuli in a predetermined "random" order. Signs were placed at either end of a flat straight stretch of 2-lane (1,620 m) paved, undivided highway with wide shoulders. Task: subject drove towards signs and was required to make 2 verbal responses during each trial (a) classify sign (regulatory or warning) (b) indicate meaning as soon as it was legible. Measure: classification distance, identification distance (at 2 speeds, 30 mph and 50 mph.
Exp't #2 designed to measure classification distance and identification distance of miniature traffic signs (1/3 size) for subjects driving 1/3 of the fast speed used in exp't #1 (17 mph). Subjects: N=16. Method: same message used in exp't #1, procedure a replication of study (with course down-sized accordingly). Task and Measures: same as in exp't #1.
Exp't #3 purpose was to determine verbal reaction time required to classify and identify traffic signs of different types. Method: stimuli were 26 slides of traffic signs rear-projected onto a screen, 15 of these signs were same as used in exp'ts 1 and 2. Half of subjects performed classification task first, half did identification task first. Tasks: classification task = subject responded "yes" if regulatory or warning sign, "no" if information sign. identification task = give verbal meaning as quickly as possible. Measures: verbal response times.

Results
- symbols were identified better than verbal signs in road studies but not in lab
- performance better on warning signs than for regulatory for all 3 exp'ts
- classification faster than identification for all 3 experiments
- similar information about the relative adequacy of warning and regulatory traffic signs can be obtained from all three methods

STATIC SIGNS LAB COMPREHENSION

Dewar, R.E. and Ells, J.G., The Semantic Differential as an Index of Traffic Sign Perception and Comprehension, Human Factors, 1977, 19(2), 183-189.

Two experiments were conducted to investigate the relationships between (i) semantic differential scores and comprehension, and (ii) between semantic differential scores and glance legibility.
Exp't #1 Subjects: N = 31. Stimuli: colour slides of 20 symbolic traffic signs. Method: each stimulus was presented on a screen for 15 seconds. After all 20 had been shown, each was presented again. Tasks: first presentation - subjects recorded the meaning of each sign on an answer sheet; second presentation - subjects were to rate each sign on the semantic differential scale. Measures: % correctly identified (comprehension), semantic differential scores.
Exp't #2 Subjects: N = 20. Stimuli: 20 traffic sign messages (10 verbal, 10 symbols). Method: for the first presentation, each slide was shown for a duration of 40 ms, for the second presentation, each slide was shown for 40 seconds. Tasks: first presentation - subject was to match stimulus with coloured replicas of 20 signs on cardboard in front of him. second presentation - subject was to rate each sign on semantic differential scale.

Overall Results

- semantic differential (SD) scores on all 4 factors (evaluative, activity, potency, understandability) were highly correlated with comprehension of symbolic messages.
- SD scores were unrelated to glance legibility of verbal messages
- 2 SD factors (evaluative, understandability) did correlate with glance legibility of symbolic messages

STATIC SIGNS LAB PROCESSING TIME

Dewar, R.E., Ells, T.G. and Mundy, G., Reaction Time as an Index of Traffic Sign Perception, Human Factors, 1976, 18(4), 381-392.

A series of 3 experiments were conducted in an attempt to compare laboratory reaction time measures of traffic sign perception with perception of signs in an actual driving situation.
Exp't #1 16 subjects (8 males, 8 females) participated. Method: slides of 20 traffic sign messages were presented on a screen. Each subject performed two tasks - classification and identification - counterbalanced between subjects for order. Tasks: classification (subject responded "yes" if sign was regulatory of warning); identification (subject responded with verbal meaning as quickly as possible). Measures: verbal reaction time.
Exp't #2 objective was to study the effect of a secondary task on sign perception. Subjects: N = 16. Stimuli: same as in exp't #1 (only larger signs used). Tasks: subjects were to classify and identify signs as well as to respond to a set range of visually presented numbers by pressing a response key. Measures: reaction time.
Exp't #3: objective was to study the effect of a loading task designed to divide the subject's attention. Subjects: N = 16. Stimuli: same as in exp't #1, however signs were front-projected onto a colored motion picture of a two-lane rural highway. Task: subjects were to respond to signs (classify and identify) as well as adjust knob to maintain a speedometer at 80 km/h. Measures: reaction time.

Overall Results

Similar trends were found in the three experiments.
- reaction times were shorter for the classification task than for identification
- reaction times were shorter for warning than for regulatory signs
- reaction times were shorter for verbal than for symbolic messages
- comparison of the above reaction times with field study data (legibility distance) revealed significant correlations

STATIC SIGNS LAB PROCESSING TIME

Ells, J. G. and Dewar, R.E., Rapid Comprehension of Verbal and Symbolic Traffic Sign Messages. Human Factors, 1979, 21(2), 161-168.

Two experiments in which the "same-different" reaction time was used.
Exp't #1 Subjects: N = 12. Stimuli: 16 coloured slides (sign type = 8 warning , 8 regulatory) 4 of the messages for each sign type were verbal and 4 were symbolic. Method: experimental session consisted of 112 trials. Task: subject was to respond "yes" if meaning of the visually presented traffic sign matched the vocally presented antecedent referent message, and respond "no" if there was no match. The verbal response terminated the stimulus and the clock counter. Measures: reaction time, errors.
Exp't #2 Subjects: N = 24. Stimuli: 28 coloured slides consisting of a verbal and symbolic message for each of 14 messages. Subjects were tested in a vision tunnel and were tested on all signs under both visually degraded and nondegraded conditions. (7 of the 14 message were warning signs and the other 7 were regulatory signs). Half of the subjects viewed only regulatory signs and the others viewed only warning signs. Task: subjects were to respond "yes" if antecedent message and referent message were the same and "no" if they were different. Measures: response times.

Overall Results

- reaction times for symbolic signs were shorter than for those signs with verbal messages
- visually degrading the sign resulted in a greater decrement in performance for verbal than for symbolic signs.
- reaction time correlated significantly with a previously obtained measure of sign legibility taken from a field study

STATIC SIGNS

ANNOTATED BIBLIOGRAPHY

LEGIBILITY DISTANCE
COMPREHENSION
CONSPICUITY
PROCESSING TIME

Forbes, T.W., Snyder, T.E., and Pain R.F. A Study of Traffic Sign Requirements. II Annotated Bibliography, Dept. of Psychology, Michigan State University, August, 1964.

This annotated bibliography includes abstracts under the following headings: (1) Methodology; Combinations of Factors; Use and Understanding of Traffic Signs; (2) Detectability and Visibility; (3) Legibility; (4) Effects of Acuity and Dynamic Visual Acuity; (5) Attention Factors.

STATIC SIGNS LAB LEGIBILITY DISTANCE

Gordon, D.A. Legibility of Highway Guide Signs, TRB/TRR, 1982, 855, 1-6.

50 subjects participated in a study of cardinal-directional sign legibility and guide sign legibility. Method: Data was collected on a vision testing alley with scaled-down replicas (1/45) of highway signs. Four different viewing distances were set at 25 ft., 30 ft., 40 ft., and 50 ft. Three tests were taken, (1) Snellen E (2) block and capital/lower case (CLC) directional charts (3) highway sign replicas. Two groups for test #3: experimental group (N = 25) viewed hignway signs with CLC, control group (N = 25) viewed signs with block letters. Task: subjects were to read out loud signs at 4 different viewing distances. Measures: correct response, legibility distance.

Results

- cardinal direction signs showed good legibility
- route numbers had poorest legibility
- CLC cardinal-direction signs were seen 10% farther away than block
- use of CLC lettering can increase legibility without requiring appreciably more space
- feasibility of performing sign legibility research on small scale replicas supported
- minimum letter height of .85 cm recommended to achieve adequate subject testing distance

STATIC SIGNS | LAB | COMPREHENSION PROCESSING TIME

Green, P. and Pew, R.W. Evaluating Pictographic Symbols: An Automotive Approach, Human Factors, 1978, 20(1), 103-114.

50 subjects (25 males, 25 females) participated. Method: 19 pictographic symbols used to label automobile controls along with their official names were used as test stimuli. 30 subjects participated in one day of testing, while the other 20 participated in 3 days of testing. Five different experimental tasks were used: Task #1: familiarity task (subjects indicate whether or not they had seen test symbols before), Task #2: Association task (subjects point to the appropriate symbol for the item described), Task #3: Rating task (rate how communicative symbol was), Task #4: Paired-Associate (PA) learning task (subjects paired labels with symbols), Task #5: Response time task (a pictograph and label were presented on a screen, subjects responded "same" or "different" by pressing appropriate button). Measures: T1: familiarity score. T2: correct responses. T3: ratings. T4: correct response for PA, T5: response time.

Results

(1) Sex and extent of one's technical background appeared to affect ability to identify unknown symbols at first glance.
(2) Reaction time in a discrimination task is influenced by individual differences and decreases markedly with learning.
(3) The strength of the association between symbols and labels is not necessarily related to how easily those symbols are learned and is only weakly related to how rapidly those symbols are discriminated once learned.
(4) If symbols are to be easily learned, labels should be brief and understood by the public.
(5) Ratings of communicativeness are correlated with associative strength, and the former can be used as an estimate of the latter.

STATIC SIGNS LAB CONSPICUITY
CAR-DRIVER RESPONSE
COMPREHENSION

Hulbert, S. and Beers, J. Wrong way driving: off-ramp studies, Highway Research Records, 1966, 122, 35-49.

Two phase study. Phase I: 4 different DO NOT ENTER signs were evaluated (1. white on black, 2. black on white, 3. white on red, 4. white (letters and bar) on red). Autopip (see p. 36) used to indicate 6 location cues. "Drive" on mountain road and residential area. 81 subjects tested on 4 signs (each subject saw only 1 sign). Measures: questionnaire and verbal interview; observed reactions, response measurements.

Results

- most subjects could report seeing DO NO ENTER
- response measurements: deceleration time and distance; braking time and distance
- red signs more effective both more often and earlier than black on white signs (lane-line arrowheads) preliminary study -- arrows difficult to detect

Phase II 8 films of "off ramp" 3 primary signs: (a) black letters on white, (b) white letters and bar on red, (c) black letters on white and red DANGER. secondary signs (a) WRONG WAY GO BACK (white on red); (b) STOP GO BACK (white on red); (c) GO BACK YOU ARE GOING WRONG WAY (white on red); (d) sign (b) with DANGER above (e) sign (a) combined with DANGER

Procedure 243 drivers, wrong way ramp film seen as a part of a series of films. An oscillograph recorded driving speed, accelerator pedal patterns, brake pedal patterns, steering patterns

Interpretation

- red primary sign is more visible than black and white
- addition of red DANGER sign significantly shortens reaction time to black and white sign

STATIC SIGNS LAB COMPREHENSION

Hulbert, S., Beers, J. and Fowler, P. Motorist's Understanding of Traffic Control Devices. AAA Foundation for Traffic Safety, March 1979.

3,164 motorists from all over the United States participated in the test program of driver understanding of different traffic control devices. Task: Subjects viewed a 16 mm colour film showing 23 different scenes of 16 different control signs, signals and markings. Subjects were asked questions concerning a scene through the soundtrack and were to respond by selecting a multiple choice answer presented on the prepared questionnaire. Measure: first-choice response. (Responses were analyzed by region of the country in which the drivers did most of their driving, as well as by age and sex of the driver.)

Results: No control device was understood by all drivers. Even the best understood devices were not well understood by 5 to 10 % of the drivers. Symbol signs were generally understood better than signals or pavement markings. Solid single yellow pavement markings and "orange" background signs were not well understood. Older drivers (> 49 years) showed less accurate comprehension of traffic controls than younger drivers.

STATIC SIGNS LAB LEGIBILITY DISTANCE

Jacobs, R.J., Johnston, A.W. and Cole, B.L. Visibility of alphabetic and symbolic traffic signs, Australian Road Research, 1975, 5(7), 68-86.

Paired alphabetic and symbolic displays of 16 familiar traffic sign messages were examined in two complementary experiments,
Exp't #1 10 subjects with normal visual acuity viewed the signs which were presented with and without their standard borders. Method: signs were reproduced as black and white transparencies in 27 sizes, representing standard sized signs viewed at distances from 22 m. to 385 m. Test signs were presented on a screen (in random order) with a viewing time of 600 ms. Task: subject responded to each slide by stating his perception of the meaning of the sign. Measure: lebilility distance.
Results
- 50 % threshold legibility distance of alphabetic signs is twice that of symbolic signs (142 m. vs 270 m.)
- presence or absence of sign borders has no effect on legibility distance

Exp't #2 5 subjects with normal vision participated in a study designed to investigate the effects of blurred vision on the legibility of alphabetic and symbolic signs. Method: same method as above, subjects viewed signs (with borders) through low power positive lenses. These lenses defocus the eye and reduce visual acuity. Task: same as above.
Measure: same as above.
Results
- reduced visual acuity reduces legibility in a proportional fashion
- shape coding on signs does not enhance their legibility among a set of signs
- sign size required for 0.95 probability of correct identification is approx. 1.7 times larger than size for 0.50 probability of correct identification

STATIC SIGNS FIELD CONSPICUITY

Johansson, G. and Backlund, F. Drivers and road signs. Ergonomics, 1970, 13, 749-759.

This study was designed to investigate the function of the road sign system as an information channel for car drivers. A stretch of road on a main highway was selected as the testing site. One of 5 experimental road signs was set up so that drivers of northbound vehicles would have a clear view of it when they were about 400 m away. A police barrier was set up further down the road (700 m.). 5,000 drivers were stopped by police and questioned by researchers. Task: answer short questionnaire. Measure: recall.

Results: The overall probability of a road sign being recalled was not greater than 0.5. The different signs studied form a scale recording probability of perception extending from a low group with probability of being perceived of 0.25 to a group with probability of 0.6 to 0.75. Rank order of signs (high probability to low) were consistent between occassions (i.e. speed sign, police control, breaks in road surface, other danger, pedestrian crossing).

STATIC SIGNS REVIEW LEGIBILITY DISTANCE

Johnston, A.W., Cole, B.L., Jacobs, R.J. and Gibson, A.J. Visibility of Traffic Control Devices: Catering for the Real Observer. Ergonomics, 1976, 19(5), 591-609

This paper reviews the visual characteristics of the 'real' observer using both conventional methods of assessment and an information theory approach. Recent research by the authors and their colleagues on the acquisition of information by observers with normal and degraded vision shows that present scales for specifying visual performance do not adequately represent the visual requirements of many practical tasks. Topics covered include visual acuity and legibility distance, signal recognition and effects of blurred vision on legend legibility.

STATIC SIGNS LAB PROCESSING TIME
LEGIBILITY DISTANCE
COMPREHENSION

Mackett-Stout, J. and Dewar, R.E. Evaluation of Symbolic Public Information Signs, Human Factors, 1981, 23(2), 139-151.

Eight public information messages were evaluated in a series of four experiments.
Exp't #1 Subjects: N = 16. Stimuli: black and white slides were made of 32 symbolic public information signs (4 versions of 8 messages). A display of the 32 symbols was prepared, with symbols arranged such that 4 versions of each message were grouped together, the meaning printed above and each version labeled 1 to 4. Method: stimuli were rear projected onto a screen and were presented for 40 ms slightly out of focus. Subjects were permitted to advance slides at their own rate. Task: match the stimuli with symbols on display and record responses on an answer sheet. Measure: glance legibility.
Exp't #2 Subjects: N = 32. Stimuli: same as Exp't #1 only mounted photographs (.5 cm X .5 cm). Method: subject sat at desk with viewing box at eye level. Each stimulus was advanced toward subject at a rate of 4 cm/sec. Task: identify sign as one of eight referents. Measure: legibility distance.
Exp't #1 Subjects: N = 148. Stimuli: same as in exp't #1. Method: subjects were divided into 4 groups. Slides were projected onto a screen. Each group saw one version of each referent. Task: subjects were required to indicate on an answer sheet what they believed the symbol meant. Measure: comprehension.
Exp't #4 Subjects: N = 40, subjects divided into 4 groups for testing. Stimuli: same as for Exp't #1. Method: each of 32 symbols was presented on a screen, with the appropriate meaning given orally. Task: subjects were to rate each symbol for clarity of meaning (1 = very unclear, 6 = very clear). Measure: preference.

General Results

- preference measure was significantly correlated with comprehension and legibility distance
- comprehension correlated significantly with legibility distance
- glance legibility did not correlate significantly with any of the other three measures

STATIC SIGNS LAB PROCESSING TIME

McNees, R.W. and Messer C.J. Reading Time and Accuracy of Response to Simulated Urban Freeway Guide Signs, Texas Transportation Institute, January 1982.

This study was designed to determine the time required by motorists to read urban freeway guide signs and the accuracy with which they read the signs. Two similar laboratory studies were conducted. Subjects: for the first study, N = 87, for the second, N = 70. Stimuli: slides made of guide signs. Variables included (a) the number of panels per overhead sign structure (2, 3, 4, 5 panels), (b) the number of "units" of information on each panel (2, 4, 6, 8, 10 units), (c) the display time for subjects to "read" the signs (2.5, 4, and 6 seconds). Method: subjects were asked to follow a hypothetical route through an unfamiliar city based on: (1) navigational directions provided by a schematic map of the area and (2) simulated guide signing presented by 35 mm slides at 22 locations along the route. For study 1, the first "trip" display time = 6 sec., for the second "trip" display time = 4 sec., and for study 2 display time = 2.5 seconds. Tasks: subjects were to select the sign panel which provided information to the final destination from the set of panels on the sign bridge. Subjects responded by selecting the corresponding response button. There were 5 buttons corresponding to the 5 possible panel positions. Measures: reaction time, accuracy level.

Results

- the optimum accuracy level was about 6 units of information per panel
- when the information level was less than 16 units per sign bridge , 100 % of subjects could read the signs acceptably; 16-30 units, 51% could read sign acceptably; 31-50 units, only 33% of subjects could read sign acceptably
- route selection accuracy decreases as the number of route choices increase
- time required to read a sign also increases with the number of route choices and total information on the sign
- on a large sign (3 or more panels) the information content should not exceed 16 units of information per sign bridge

STATIC SIGNS LAB . PROCESSING TIME

Noble, M. and Sanders, A.F. Searching for Traffic Signals while Engaged in Compensatory Tracking. Human Factors, 1980, 22(1), 89-102.

Eight subjects participated in an experiment to study the effects of performing a search task and a tracking task simultaneously. Search Task: The stimuli for the search task were 12 circular and 12 triangular traffic-like signs. Stimuli were presented in two clusters one on either side of the screen. Subjects were instructed to search for only those stimuli which appeared on the bottom row of the screen. The task then was to search through both clusters and indicate the total number of relevant signals present by pressing the appropriate button on the response panel. Measure: reaction time. Tracking Task: Subjects tracked a path using a white cursor presented in the middle of the screen. Measure: error score. Five independent variables were manipulated: total number of relevant signals in that display, density of signals, colour conditions, and track plus visual search versus visual search alone.

Results: Performance degradation was found for both tasks when tracking and visual search were performed concurrently. Examination of main effects and interactions suggest that, except when colour was a perfect cue, visual search was a sequential process.

STATIC SIGNS FIELD COMPREHENSION
CONSPICUITY
DRIVER-CAR RESPONSE

Seymour, W.M., Deen, R.C. and Havens, J.H Traffic Control for Maintenance on High Speed Highways. Final Report. Research Report 327, Kentucky Dept. of Highways, Division of Research, 1972.

Field Observation: Over 15,000 observations were made at actual lane closures on Interstate highways to compare effectiveness of yellow and orange signs. Test signs were identical except for colour. New yellow signs were hung over the contractor's signs and traffic was observed for one hour. Then the new orange signs replaced the yellow and traffic was observed for an hour. Measures: driver behaviour (spot speed checks, traffic conflicts, merging manoeuvers, and turn signal application). Driver Interviews: 62 drivers were interviewed after they passed through a lane closure. Sign colours were alternated so that drivers could make comparisons. Measures: recall, comprehension, preference.

Results: Driver obedience improved when new signs of either colour were used. Orange signs were slightly more effective than yellow signs in reducing traffic conflicts and merges near traffic cones. Driver preference supported orange signs more strongly.

STATIC SIGNS FIELD LEGIBILITY DISTANCE

Sivak, M., Olson, P.L. and Pastalon, L.A. Effect of Driver's Age on Nighttime Legibility of Highway Signs, Human Factors, 1981, 23(1), 59-64.

Nighttime driving study (closed-course) investigated the effect of driver's age on legibility of highway signs. Subjects: 12 younger subjects (18-24) and 12 older subjects (62-74), subjects were closely matched on high-luminance far acuity. Stimuli: a number of test signs were developed using 7 standard highway colours and reflective materials used to mark right or left-oriented E. Method: data was collected from 3 subjects at a time (1 driver, 2 passengers). Task: subjects were to press response button once for regular E and twice for reversed E. Measure: legibility distance.

Results

- legibility distances for older subjects were 65 - 77 % of those of the younger subjects.

STATIC SIGNS FIELD CONSPICUITY

Summala, H. and Naatanen, R. Perception of highway traffic signs and motivation. Journal of Safety Research 1974, 6(4), 150-154.

Nine subjects, individually drove 257 km. on a 2 lane road with heavy traffic. Task: Each subject was asked to drive carefully and name out loud all of the traffic signs on the route. There was a total of 581 traffic signs. Measure: number of traffic signs named (comprehension).

Results: In the urban areas only 3 % of the signs went unreported, while only 2 % of the signs in the rural area were not reported. Authors suggest that highly motivated drivers can report practically every sign in a highway driving situation.

STATIC SIGNS | LAB | LEGIBILITY DISTANCE
PROCESSING TIME

Testin, F.J. and Dewar, R.E. Divided Attention in a Reaction Time Index of Traffic Sign Perception, Ergonomics, 1981, 24(2), 111-124.

48 subjects participated in an experiment designed to study division of attention by requiring subjects to identify traffic signs while performing a secondary task. Method: Three types of 35 mm slides were used (i) coloured slides of 16 regulatory signs (4 regulatory verbal, 4 regulatory symbolic, 4 warning verbal, 4 warning symbolic) used by Dewar and Ells (1974). (ii) 16 typewritten messages corresponding to 16 slides (iii) matrices of 4 numbers and blank slides (loading task stimuli), subject responded to slide by pressing appropriate button "same", "diff", "yes", "no". Tasks: subjects viewed slides of messages followed by traffic signs and indicated using a response button whether the message and the sign were the same or different. Subjects performed this task alone or in combination with one of three loading tasks; a detection task (did a superimposed slide contain numbers), a identification task (did the superimposed slide contain the number 7) and a memory task (did the superimposed slide contain the same number as a previous slide).
Measures: reaction times for correct responses (correlation coefficients were calculated between average reaction time and legibility distance determined by Dewar and Ells (1974).

Results

- warning signs are more legible than regulatory signs and symbolic signs are more legiblethan verbal signs
- validity of a reaction time index of traffic sign perception appears not to be increased by the loading tasks used in this study

STATIC SIGNS LAB CONSPICUITY

Woltman, H.L., Stanton, R.H., and Stearns R.A. Information Sign Color Evaluation Employing a Video Presentation, Traffic Control Materials Division/3M Audio Visual Communication, St. Paul, Minn. 55101, January 1984.

313 subjects of various ages participated in a laboratory study evaluating the effectiveness of green nighttime sign color. Method: subjects viewed 12 video presentations of 6 identical pairs of nighttime highway scenes in which only the guide sign color varied (white on green, white on black). Viewing times for scenes were set at 3.5 seconds. (Nighttime slide scenes and miniature signs were used in the creation of the videotaped scenes). Task: subjects were to count the number of signs they could find in each scene. Measure: number of signs missed by color (error rate).

Results

- greater scene complexity and driver age contributed to an increased error rate for both color combinations
- for all scenes combined, accuracy (identifying) was 3.2 times greater for the white on green signs than for the white on black signs

DYNAMIC SIGNS QUESTIONNAIRE COMPREHENSION

Dudek, C.L., Messer,C.J. and Jones, H.B. Study of Design Consideration for Real-Time Freeway Information Systems. Highway Research Record, #363, p. 1-14, 1971.

505 licensed drivers were surveyed from various business organizations in the cities of Houston and Dallas. Task: to respond to multiple choice questions of questionnaire (number of items in questionnaire not specified) Measures: driver attitudes and preference for changeable message signs, probable reaction to real-time freeway information.

Results

(1) participating motorists indicated desire for additional traffic information not currently provided by passive signs
(2) changeable message signs and commercial radio preferred over telephone and television services
(3) drivers preferred to receive information about traffic conditions before entering freeway so they could divert to an alternative route
(4) types of information preferred were the location and length of congested area and degree of congestion
(5) quantitative description of travel time and travel speed were least preferred

DYNAMIC SIGNS FIELD

Dudek, C.L., Stockton, W.R. and Hatcher D.R. Real-time Freeway-to-Freeway Diversion -- The San Antonio Experience, Texas Transportation Institute, January 1982 (draft).

Study to evaluate the effectiveness of a low-cost motorist information diversion system in San Antonio, Texas. Method: assessing change in traffic volumes on freeway, interchange ramps [evaluation of sign messages: 19 subjects drove route, when subject within legibility distance of CMS, flash cards were shown with messages typical of what would be shown on CMS, 84% chose correct route, modifications made, 10 more subjects tested -- all chose correct route] Measures: traffic counters, sampled when (1) CMS used (2) CMS not used (3) normal volumes for same day of week (CMS used during 15 actual accidents and 7 non-accident incidents)

Results

- overall low diversion rate
- diversion ramp was too close to the final destinations of divertable drivers
- drivers were using routes other than diversion route when they saw messages on CMS. Diversions on other routes were not evaluated

DYNAMIC SIGNS FIELD DRIVER RESPONSE

Dudek, C.L., Weaver, G.D., Hatcher, D.R. and Richards, S.H. Field Evaluation of Messages for Real-Time Diversion of Freeway Traffic for Special Events, Transportation Research Board Record #682, 37-45.

3 special event route diversion studies conducted in Dallas to evaluate 14 primary candidate messages (events: July 4, 1976, Oct. 9, 1976 state fair, January 1, 1977, Cotton Bowl game) Measures: comprehensive license-plate origin-destination technique.

Results

(1) all matrix-signs evaluated greatly influenced route diversion to alternate route
(2) before events that have fixed starting times, all candidate messages are equally effective in diverting traffic
(3) before events for which arrival times are not set, traffic-state descriptors and specific temporal information messages produce highest percentage of diversion
(4) display of information that alternative route is guide marked produces no increases in diversion rates over the base message
(5) only one changeable message sign necessary for effective diversion

DYNAMIC SIGNS FIELD CAR-DRIVER RESPONSE COMPREHENSION

Hanscom, F.R. Effectiveness of Changeable Message Signing at Freeway Construction Site Lane Closures, Transportation Research Record, #844, 35-41, 1981.

Right- and Left-lane closures situations were studied and observations were made under day and night conditions (more than 196,500 vehicles), also a driver questionnaire was answered by 489 drivers. Measures: unobtrusive: vehicle speed, lane distribution (30 minute data-collection intervals). questionnaire: sign detection, comprehension, interpretation.
Message Conditions: (1) speed and closure advisory (2) speed and merge advisory (3) merge and closure advisory (4) closure advisory.

Results

- before and after studies of baseline (no CMS) versus application demonstrated increased advance preparatory lane-change activity, significantly fewer late exits, smoother lane-change profiles, reduced speeds at lane closure point
(2) most preferable CMS location 1.2 km in advance of closure
(3) 3 devices tested (i) one-line bulb matrix (ii) 2-line rotating drum (iii) 3-line bulb matrix
- 3-line bulb matrix tended to produce more advance lane-changes
- no difference in hazardous late exit manoeuvres observed between types
- all three equally effective in eliciting speed reductions at entrance of lane closure
- drivers preferred 3-line bulb matrix
(4) drivers preferred speed and lane closure message combination
(5) study recommends CMS application as a supplement to standard devices but not as a replacement for arrow-boards

DYNAMIC SIGNS QUESTIONNAIRE COMPREHENSION

Heathington, K.W., Worrall, R.D. and Hoff, G.C. An Analysis of Driver Preference for Alternative Visual Information Displays, Highway Research Board Record #303, 1-11, 1970.

732 drivers (Cook County, Illinois) responded to a structured questionnaire to obtain responses for paired comparison evaluation traffic descriptors: 7 descriptors for heavy congestion, 6 for moderate and 6 for no congestion. Task: Each descriptor at each level of congestion paired with all other descriptors at that level, subjects told that the signs (descriptors) in each pair described the same traffic conditions on an expressway. Subjects asked to indicate whether he preferred sign A or sign B as a form of traffic information. Measures: Driver Preference.

Results

(1) for all levels of congestion, traffic information was preferred to no information about traffic conditions
(2) for level of heavy congestion information about accident causing heavy description preferred descriptor, speed descriptor ranked second
(3) speed descriptor ranked first for "moderate" and "no" congestion levels
(4) the two quantitative descriptors (delay and travel time) not desired by respondents

DYNAMIC SIGNS FIELD LEGIBILITY DISTANCE

Hodge, A.R. and Rutley, K.S. A Comparison of Changeable Message Signals for Motorways, Report # PB-30134/5ST, 1978.

Evaluated 5 types of changeable message signs under day and night conditions. Method: 25 subjects, driven one at a time, as front seat passengers, towards signals from 600 m away (speed = 15-20 km/h). Signals under test arranged in line abreast facing an unobstructed sight line. One signal illuminated on each run. Task: identify legend on sign as soon as it could be read clearly. Measure: recognition distance (measured by meter in car). 38 readings obtained for each subject using selection of legends (numerals 40, 60, 100 or arrows left and right) on five signs. (also recorded time of day, weather, light levels, sun direction, subject details).

Results

(1) in daylight the recognition distances of numerals tested on matrix and roller-blind signals were similar, but at night latter were greater
(2) in good visibility, the better of the two roller-blind signals had the longest recognition distance for numerals
(3) arrows on the better of the two roller-blind signs had greater recognition distance than those on matrix signs
(4) in reduced visibility, matrix signals had better recognition distance
(5) in most cases, signs had a recognition distance adequate for motorway speeds
(6) overall, internally illuminated matrix sign with white-on-black legends recommended as being most suitable for use on motorways

DYNAMIC SIGNS LAB COMPREHENSION

Huchingson, D. and Dudek, C.L. Delay, Time Saved and Travel Time Information for Freeway Traffic Management, Transportation Research Record #722, p. 36-39.

5 studies of freeway motorists' opinions were conducted to determine preferences and reported behaviour with respect to hypothetical displayed messages about time delays.
Study 1: levels of delay time, Subjects: N = 464. Method: 7 cards each (type of incident, delay period). Task: check "yes" stay on freeway or "no" get off. Measure: diversion decision. Result: 50% divert if 15-20 min delay.
Study 2: levels of time saved. Subjects: N = 127. Method: same as above. Results: most would divert if 5 - 10 minutes saved.
Study 3: major and minor accident messages. Subjects: 40 receive "major" message, 20 receive "minor" message. Task: indicate number of minutes of delay message implied.
Results: minor = approx. 12 min. delay; major = approx. 22 min. delay.
Study 4: meaning of delay. Subjects: N = 40. Method: told accident ahead with 30 minute delay. Task: check on a 5-point Likert scale agreement with each of 5 interpretations. Measure: interpretation. Results: delay = something relative to unusual conditions at that time of day.
Study 5: modes of presenting temporal information: survey of 70 drivers. Method: presented subject with scenario and 3 different sign messages. Task: indicate which of three messages would convince subject to divert off freeway.
Measure: preference. Results: 88 % thought messages said the same thing.

DYNAMIC SIGNS LAB CAR-DRIVER RESPONSE COMPREHENSION

Hulbert, S. and Beers, J. Research Development of the Changeable-Message Concept for Freeway Traffic Control, Highway Research Board Special Report, #129, 1971, 24-25.

Used three techniques (UCLA sign tester, the Preview house, and the UCLA Driving Simulation Laboratory) to study the types of messages appropriate for drivers as they approach congestion. Method: not given. Measures: driver preference, driver response (to 80 specific messages)

Results

(1) subjects indicated a preference for sign-message system over an audio system
(2) in terms of driver response either sign or audio system will favourably modify behaviour
(3) content of messages should be short and simple
(4) audio messages should be limited to 2 or 3 short sentences

DYNAMIC SIGNS LAB COMPREHENSION

Knapp, B.G. and Pain R.F. Human Factors Considerations in Arrow-Board Design and Operation, TRB/TRR #703, p. 1-8, 1979.

Two lab studies made use of film clips to evaluate comprehension of various arrow-board configurations. Study 1 20 subjects participated. Method: subjects shown 9 film clips, each one represented different modes of arrow-boards in combination with placement on lane or on shoulder (each clip shown twice, random order) Task: after each clip, subject chooses one of 4 responses indicating meaning of arrow, and subject also gives a rating of his confidence in response Measure: comprehension, confidence. Result: Flashing arrow in directional connotes lane-closure ahead.

Study 2: 109 subjects participated. Method: 6 film clips, 2 presented simultaneously. Task: subject was to indicate, by checking either left or right, which mode of the two presented in each film clip best conveyed the meaning of lane closure. Measure: driver preference. Result: a blinking arrow and a sequential chevron were superior to a sequential arrow.

DYNAMIC SIGNS LAB COMPREHENSION

Mounce, J.M., Messer, C.J., Huchingson, R.D. and Dudek, C.L. Bulb-Loss Effects on Message Readability of Motorist-Information Matrix Signs. TRB/TRR #643, p. 40-45, 1977.

93 subjects gave written responses to slide presentations of matrix signs with varying amounts of bulb loss. Method: independent variable = characters per word and size of matrix, location of bulb failure, percentage of bulb failure in 10 % increments, single word presentation rate of 3 secs / word. Each subject saw 100 words. Task: legibly write the word or numeral within 10 secs. after presentation. Measure: percentage of correct responses (two states -- familiar (seen whole word before) and unfamiliar)

Results

- at all levels of bulb loss unfamiliar state elicited poorer recognition than familiar state
- length of word had no systematic relationship to percentage correct response for familiar state
- for the unfamiliar state, longer words were somewhat more difficult to recognize than shorter words
- for familiar driver, bulb losses of 45 and 30% (corresponding to 85th and 95th percentile design levels) are tolerable before deterioration reaches a point where legibility is a problem
- for unfamiliar drivers, tolerable bulb losses are 20 % and 10 % (for 85th and 95th percentile design).

DYNAMIC SIGNS QUESTIONNAIRE CAR-DRIVER RESPONSE

Richards, S.H., Stockton, W.R. and Dudek, C.L. Analysis of Driver Responses to Point Diversion for Special Events, TRB/TRR #682, p. 46-52, 1978.

In conjunction with a set of field studies (special-event route diversion), a questionnaire was designed to (a) obtain driver attitudes and reactions to route-diversion systems (b) explore trends in driver-behaviour related to diversion decision (c) identify factors that influence a driver's decision to divert. Two questionnaires were developed, one for those who had diverted in field study and one for those who had not. Questionnaires mailed to respondents one month after field study. Effective rate of return = 27% (N = 480). Measure: driver attitudes, reasons for (or for not) diverting.

Results

(1) drivers who were more familiar with the route were more likely to divert
(2) anticipated dissatisfaction with alternative route was most frequently cited deterent to diversion
(3) advisories of severe traffic problems were cited most often as the information that would have influenced those drivers who did not divert to do so.
(4) additional information for parking for special events was often requested
(5) no drivers who did divert stated that they would not follow diversion instructions in future

DYNAMIC SIGNS FIELD LEGIBILITY DISTANCE COMPREHENSION

Rutley, K.S. and Webb, P.J. Recognition distances and understanding of legends on an experimental motorway signal. TRRL Supplementary Report 500, 1979.

A prototype three-part signal was built to display information so as to advice motorists of the presence of a motorway hazard as well as its nature and location. A number of pictographs were designed for indicating the nature of the hazard. The study was conducted in two parts, designed to test signal recognition distance and pictogram understanding. Study 1: 24 subjects were driven individually towards the signal situated on the test track. 12 different legend combinations were displayed. Subjects were familiarized with the 12 pictograms prior to testing. Task: subject was to identify the three-part legend as soon as possible. Measure: recognition distance. Study 2: 35 subjects who had not previously seen the pictograms participated. Groups of 8 subjects viewed the signal from a bus parked within easy viewing distance. Initially each symbol was displayed for 7 seconds, later the trial was repeated with the pictogram displayed for 30 seconds. Task: subjects were to write down their interpretations of each symbol. Measure: correct response (comprehension).

Results: The tests showed that the top advisory panel and the 'distance ahead' panel could be recognized from distances adequate for motorway speeds. Some pictographs had to be enlarged by 60% so as to be recognized at an adequate distance. Pictograms which were similar to those currently in use were well understood. New designs were understood by between 40 and 100% of subjects.

DYNAMIC SIGNS LAB LEGIBILITY DISTANCE PROCESSING TIME

Snyder, H.L. and Taylor, G.B. The Sensitivity of Response Measures of Alphanumeric Legibility to Variations in Dot Matrix Display Parameters. Human Factors, 21(4), 457-471, 1979.

An evaluation of the sensitivity of four observer response measures (recognition accuracy, response time, tachistoscopic recognition accuracy, and threshold visibility) Method: alphanumeric characters were presented to six subjects in non-contextual form on a variable-parameter CRT. Display parameters investigated included 4 character sizes, 3 luminance levels, and 7 viewing distances (each subject received every level). Measures: (1) accuracy and response time: each subject was presented 18 randomly chosen alphanumeric characters under each experimental condition (2) tachistoscopic recognition: each subject was given combinations of 3 luminances and 4 character sizes. Exposure times were 16.67, 33.33, 50 ms (12 random characters presented at each exposure time) (3) threshold visibility: data collected for accuracy response measure were also used to determine the 50% and 85% threshold viewing distances for all 12 C X L combinations. Task: press hand-held button to start trial, press button as soon as recognize character, then key in response.

Results

(1) response accuracy was found to be the most sensitive measure
(2) character size, character luminance, and viewing distances proved to have consistent and significant effects at viewing distances greater than 2.35 m
(3) by increasing luminance, response accuracy can generally be increased
(4) luminance modulation should be at least 0.78 to obtain adequate legibility (85% accuracy), while characters should subtend at 10.8 min. of visual angle at the eye

DYNAMIC SIGNS REVIEW/FIELD CAR-DRIVER RESPONSE

Stephens, B.W. Some Principles for Communicating with Drivers Through the Use of Variable-Message Displays, Highway Research Board Special Report #129, 1979, p. 2-6.

Review of general principles. Mention of extended analysis of 1966 fog study conducted by California Division of Highways: making use of Changeable Message Signs. Method: field study. Measure: coefficient of variation of speeds (inverse of traffic speed stability)

Conclusion

- lowering posted speed a little below nonsigned 'natural speeds' can improv traffic stability, lowering posted speed too much will redu stability

DYNAMIC SIGNS

FIELD/QUESTIONNAIRE CAR-DRIVER RESPONSE COMPREHENSION

Stockton, W.R., Dudek, C.L., Fambro, D.B. and Messer, C.J. Evaluation of a Changeable Message Sign System on the Inbound Gulf Freeway, Texas Transportation Institute, Research Report 200-IF, 1975.

Two techniques were used to evaluate changeable message signs (1) changes in traffic flow parameters during periods of selected message display, ("LANE BLOCKED", "SLOW TRAFFIC", "OK") to assess voluntary diversion, data was also collected after signs were removed; (2) questionnaires were mailed to drivers observed passing the signs during certain messages. Measures: (1) traffic volumes (2) comprehension, driver preference.

Results

- changeable message signs found to be effective in providing information to drivers

DYNAMIC SIGNS FIELD QUESTIONNAIRE CAR-DRIVER RESPONSE

Turner, J.M., Dudek, C.L. and Carvell, J.D. Real-Time Diversion of Freeway Traffic During Maintenance Operations. TRB/TRR #683, 1978, p. 8-10.

Messages developed in previous studies were displayed in actual field operation to determine relative effectiveness of each message. Measures: (1) change in diversion rates (flow rates of 5 min. duration) from natural diversion (nonmanagement) to diversion because of information signing (management). (2) questionnaires distributed to drivers to measure driver attitudes.

Results

- diversion was greater under managed condition than under nonmanaged
- motorists prefer to have diversion information and advisory information along route during maintenance operations

DYNAMIC SIGNS FIELD CAR-DRIVER RESPONSE

Weaver, G.D., Dudek, C.L., Hatcher, D.R. and Stockton, W.R. Approach to Real-Time Diversion of Freeway Traffic Events, TRB/TRR, 1977, 664, 57-61.

field study conducted to validate the effect of two primary candidate messages and diversion displays to be used on the day of a special event (firework display). Data Collection: comprehensive origin-destination data collected. Observers recorded (on a tape recorded) license plate numbers of vehicles passing the eight observation points. This allowed tracing of travel patterns without spot sampling and interviewing drivers.

Results

- on average 52.4% of drivers diverted when message was displayed

LANE MARKINGS
DELINEATORS

LAB/FIELD

CAR-DRIVER RESPONSE
COMPREHENSION

Allen, R.W., O'Hanlon, J.F. and McRuer, D.T. Driver's Visibility Requirements for Roadway Delineation Vol. I Effects of Contrast and Configuration on Driver Performance and Behaviour. Dept. of Transportation, FHWA/RD-77/165, Nov. 1977.

A combined theoretical/experimental approach was taken. A theory for delineation visibility and driver perceptual requirements was developed and tested in both a laboratory simulation and a closed course field study. Lab Study Subjects: N = 6. Method: a fixed-base driving simulator was used for testing purposes. The simulator had a high quality, wide angle video projection display of roadway markings. Each subject drove the entire test scenario in two or three days of concentrated testing. Each day was broken into 4 test sessions (1 1/2 hours each) separated by a rest period. At the beginning and end of each day, the 300 ft. visibility baseline configuration was tested. At the beginning and end of each session, a baseline configuration of either 300 ft. or 50 ft. visibility was administered. Experimental variables included (1) delineation configuration (dashed left and solid right; dashed left and dashed right; dashed left and blank right), (2) visibility range (35, 50, 100, 300 ft.) and vehicle speed (15, 30, 55 mph). Task: (1) driving against a random wind (2) following a winding road (3) driving an occasionally curving road as fast as was comfortable. Measures: driver behaviour, driver/vehicle system performance, "preferred speed", subjective ratings. Field Study Subjects: N = 11. Method: subjects drove highway route at night on 3 separate occasions. (1) different striping contrast levels on successive 4 mile segments (b) after delineation had been allowed to degrade (3) after highway repainted. Task: attempt to maintain constant speed (50 mph). Measures: electrophysical recordings (EEG, ECG), lane position, vehicle speed Field Study #2 Subjects: N = 11. Method: subjects drove a highway route at night under dry and wet weather conditions. Task: drive route. Measures: same as Field Study # 1.

Overall Results

- the quality of delineation under adverse visibility conditions depends on a combination of contrast and configuration (i.e. need to increase delineation contrast, increase segment to gap ratio, decrease segment cycle length, and add solid right edge line)
- suggestions are made on practical field techniques for measuring delineation contrast

LANE MARKINGS QUESTIONNAIRE COMPREHENSION

Gordon, D.A. Studies of the Road Marking Code. Federal Highway Administration, FHWA-RD-76-59, April, 1976.

Two survey studies were conducted on the communication of road marking meanings to the driver. Study 1: A total of 230 subjects answered a questionnaire designed to investigate drivers' understanding of road marking patterns and their acceptance of road marking application. The final part of the questionnaire presented traffic problems by means of a film. The film was taken from the drivers' position of view in the car. Measures: comprehension, preference. Study 2: The second study was concerned with drivers' preference for road markings as indicated by their choices of "the most logical and understandable" markings to fit a variety of common highway situations. 96 subjects answered the questionnaire. Measure: preference.

Results: Drivers favour markings which show the path of travel under adverse conditions, or which guide him in performing difficult manoeuvers. More than one quarter of the explanations were wrong of the single solid white line, single wide shite, double broken yellow, single broken yellow and single broken white markings. It was found that a broken line has a stereotyped association with permission to cross, while yellow markings are associated with hazard.

LANE MARKINGS FIELD CAR-DRIVER RESPONSE

Nedas, N.D., Balcar, G.P., and Macy, P.R. Road Markings as an Alcohol Countermeasure in Traffic Safety: A Field Test of Standard and Wide Edgelines. Presented at the Annual Meeting of the Transportation Research Board, January 14, 1981.

A closed-course field study was conducted at night to study the effect of road markings on the driving behaviour of impaired drivers. Subjects: N = 16. Method: Three BAC levels were used (0.00, 0.05, 0.08) and 4 edgeline widths (0", 4", 6", 8"). A repeated measures design was used Task: Subjects drove a test course consisting of 9.7 miles of rural two-lane roads, which included fourteen 2,000 foot sections designed to test the four edgeline widths. Measures: lane position, vehicle speed.

Results

- edgelines moderate behaviour on right side, such that there is a movement to a more central driving position with the presence of edgelines
- a significant reduction in path variability occurred with 4 inch wide edgelines (the next consistent drop was at 8 inches)
- vehicle speed did not show significant correlation with markings

LANE MARKINGS
DELINEATORS

LAB

PROCESSING TIME
COMPREHENSION

Triggs, T.J., Harris, W.G., and Fildes, B.N. Delineation Cues on Rural Roads at Night: A Laboratory-Based Study of Curve Direction Estimation. Australian Road Research Board, AIR-226-2, July 1979.

The experiment was designed to investigate several variables for two-lane highway conditions. Each subject performed under all conditions: four delineator conditions (centre line only, centre and edgelines, centre line with roadside posts, centre and edgelining with posts); three road contours (crest, flat, sag); 3 distances pre-curve (60 m, 120m, 180m); two curve directions (left, right); two instructional conditions (speed emphasis, accuracy emphasis) Subjects: N = 30. Stimuli: slides were taken from actual sites (2 examples of every combination of conditions = 144 stimuli) Method: slides were projected on a screen. Subject forwarded slides by depressing a foot switch. Two response buttons were mounted on a console, along with 5 confidence buttons lined up vertically (1 = very sure, 5 = very unsure) Task: subject to make a judgement as to the direction of the road. Measure: reaction time, errors, confidence judgements.

Results

- road-side post delineation provides effective information as to road contour, while no benefir was found for edgelining
- right-hand curves were to faster and were more easily detected than lewft-hand curves

DELINEATORS FIELD CAR-DRIVER RESPONSE

Davis, T.D. Construction Zone Safety and Delineation Study, N.J. Dept of Transportation, FHWA/NJ-83/005, February, 1983.

Five field studies were conducted in an attempt to determine delineation needs.
Exp't #1 evaluate reflectors vs. steady state lights Method: data was collected at 4 locations by 2 hidden observers. One counted all vehicles in each of 3 lanes. The other collected speed data from the barrier lane only. data collected between hours of 9 PM - 11 PM (1 hour per day, Mon. to Thurs.) Measures: lane changes, lane encroachments, vehicle speed. Results: reflectors were recommended.
Exp't #2 evaluate tall vertical panel vs. type III barricades. Method : 3 observers were used ot collect data (one to collect manoeuvre and brake light data, other 2 to collect speed and volume data). Measures: lane distribution, brake lights, speed. Results: tall vertical panels reduced lane encroachments at night.
Exp't #3: evaluate raised pavement markers as a point supplement Method: before-after design (field study with unobtrusive observers). Measures: lane distribution, lane changes and encroachments, brake application, speeds. Results: there was a reduction in lane changes and encroachments when raised markers were present.
Exp't #4 evaluate raised pavement markers as a paint replacement. Method: before-after field study. Measures: lane distribution, lane changes and encroachment, brake application, and vehicle speed. Results: raised pavement markers as a paint replacement resulted in a reduction in lane changes day and night and lane encroachments at night.

DELINEATORS LAB COMPREHENSION

Godthelp, J. and Riemersma, J.B.J. Vehicle guidance in road-work zones. Ergonomics, 1982, 25(10), 909-916.

This paper presents a functional description of delineation systems used in work zones. A laboratory experiment is discussed briefly.
Lab Study Task: subjects had to estimate the geometry of the road ahead. In one condition, one or two lights were removed in order to simulate conditions which often occur because of defects. Results: percentage of errors may reach a level of more than 70% in the case of a delineation system at eye height. Improvements could be reached by lowering the height of the delineator system.

DELINEATORS LAB LEGIBILITY DISTANCE

Pain, R.F. and Knapp, B.G. Experimental Evaluation of Markings for Barricades and Channelizing Devices. TRB/TRR, 1979, 703, 8-16.

Four laboratory experiments were conducted in an attempt to determine quantitatively the optimum design characteristics of channelizing devices. Subjects: 30 subjects participated in each experiment. Method: stimuli consisted of 4 identical "pictures" were placed together to form a square. Stimuli were presented tachistoscopically for durations of 0.4 secs or 0.8 secs. Tasks: subject was to (1) identify the type of design (horizontal, vertical, 45 degree slant, chevron) and (2) identify the shape (bar or panel). Measures: number of correct responses, confidence in Q-score (quadrant score), C-score (configuration score) and S-score (shape identification) response (1 = low; 5 = high).

Results

- optimal stripe width is 15-20 cm. (for 15 cm. or greater rail)
- optimal stripe design configuration are first vertical then horizontal
- chevrons connote directional meaning to drivers
- vertical panels elicit better performance than horizontal bars or trapezoid shapes
- little difference found between type I and type II barricades
- a tall narrow vertical panel is recommended over a shorter, wider one.

APPENDIX B : Evaluation Technique Definitions

glance legibility -- A slide of a sign or road marking is shown to a subject for a fraction of a second. The subject reports the meaning of the sign message or lane marking. The exposure time is determined in a pilot test to be of such a length that subjects respond correctly 50% of the time.

matching -- Subjects indicate which of several symbols (verbal messages could also be used) represents a specific message. This technique is useful where several similar symbols exist and might be confused.

multiple choice recognition -- Subjects are asked in a pen and paper test which of four possible meanings for a particular symbol is correct.

preference ranking -- Subjects are shown several symbols (or verbal messages) and are asked to rank them in order according to how well each conveys a particular meaning.

preference rating -- Subjects are shown several symbols (or verbal messages) and are asked to rate them on a scale of 1 to 5 according to how well each conveys a particular meaning.

reaction time -- Subjects are shown a slide of a sign or lane marking and press a response button as soon as they understand the slide contents.

recognition test -- Subjects are shown a symbol and are asked to write out its meaning.

semantic differential -- Subjects are asked to rate a symbol using several scales, for example good-bad, strange-familiar, associated with psychological meaning. Thus if good is 1 and bad 5 and the subject feels the symbol is neither good or bad he or she might rate it a 3.

APPENDIX C : STATISTICAL CONSIDERATIONS

In all the tests described in the report the sample size is not specified. The choice of sample size is a statistical matter which depends on many variables of the situation. Decisions must be made about the desired power of the tests, i.e., the significance level to be used, the actual size of the difference between alternatives to be discovered, and the presumed size of the variance of the population from which the sample subjects are drawn.

Suppose two signs are being compared. Let N be the number of subjects tested, P be the proportion of subjects preferring sign 1 and Q the proportion preferring sign 2 and Z the Z-score test statistic.

Then,

$$N = \frac{P(1-P) + Q(1-Q)}{(P-Q)^2} Z^2$$

If a 95% confidence in the results of the test is desired then the Z-score for the 0.05 level of confidence should be used.

If we are only interested in differences in preference of at least 10% (i.e. say 55% prefer one sign and 45% prefer the other) then the number of subjects we require to test at the 95% confidence level is :

$$N = \frac{.55(1 - .55) + .45(1-.45)\ (1.96)^2}{(.55 - .45)^2}$$

$$= 190 \text{ subjects}$$

Thus we would need a sample of 190 subjects to achieve statistical significance with a 95% level of confidence of detecting a difference between signs of at least 10 %. This calculation will vary depending on the test statistic being used and the number of alternatives being compared.